bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Protective coding variants in *CFH* and *PELI3* and a variant near *CTRB1* are associated with age-related macular degeneration

Erin K. Wagner[a, b,†],

Yi Yu[a,†],

Eric H. Souied[c],

Sanna Seitsonen[d],

Ilkka J. Immonen[d],

Paavo Häppölä[e],

Soumya Raychaudhuri[f,g,h,i,j],

Mark J. Daly[f,g,k],

Johanna M. Seddon[a, b, l,*]

[a] Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, MA 02111, USA
[b] Department of Ophthalmology, Tufts University School of Medicine, Boston, MA 02111, USA
[c] Hôpital Intercommunal, Hôpital Henri Mondor, Créteil Université Paris Est, Paris, France
[d] Helsinki University Eye Clinic, Helsinki, Finland
[e] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[f] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA
[g] Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA 02115, USA
[h] Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA
[i] Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA 02115, USA
[j] Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, UK
[k] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA
[l] Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, MA 02111, USA
†co-first authors
[*] Corresponding author: Johanna M. Seddon, Tufts Medical Center, 800 Washington Street, #450, Boston, MA 02111; Tel: +1 6176369000; Fax: +1 6176365844; Email: jseddon@tuftsmedicalcenter.org

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Although >20 common frequency age-related macular degeneration (AMD) alleles have been discovered with genome-wide association studies, substantial disease heritability remains unexplained. In this study we sought to identify additional variants, both common and rare, that have an association with advanced AMD. We genotyped 4,332 cases and 25,268 controls of European ancestry from three different populations using the Illumina Infinium HumanExome BeadChip. We performed meta-analyses to identify associations with common variants and performed single variant and gene-based burden tests to identify associations with rare variants. Two protective, low frequency, non-synonymous variants A307V in *PELI3* (odds ratio [OR]=0.14, P=4.3×10^{-10}) and N1050Y in *CFH* (OR=0.76, $P_{conditional}$=1.6x10^{-11}) were significantly associated with a decrease in risk of AMD. Additionally, we identified an enrichment of protective alleles in *PELI3* using a burden test (OR=0.14). The new variants have a large effect size, similar to rare mutations we reported previously in a targeted sequencing study, which remain significant in this analysis: *CFH* R1210C (OR=18.82, P=3.5×10^{-07}), *C3* K155Q (OR=3.27, P=1.5×10^{-10}), and *C9* P167S (OR=2.04, P=2.8×10^{-07}). We also identified a strong protective signal for a common variant (rs8056814) near *CTRB1* associated with a decrease in AMD risk (logistic regression: OR = 0.71, P = 1.8x10^{-07}; Firth corrected OR = 0.64, P = 9.6x10^{-11}). This study supports the involvement of both common and low frequency protective variants in AMD. It also may expand the role of the high-density lipoprotein pathway and branches of the innate immune pathway, outside that of the complement system, in the etiology of AMD.

INTRODUCTION

Advanced age-related macular degeneration (AMD) (MIM 603075) is a common, complex, chronic eye disease (1). As a leading cause of vision loss in people older than 60 years, AMD currently affects more than 1.75 million individuals in the United States. This number is expected to increase by 50% to 3 million in 2020 due to aging of the population (2). The prevalence of AMD is expanding as the

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

population ages, and therefore, the personal, societal, and economic burden is rising. The sibling recurrence-risk ratio (λ_s) for AMD is estimated to be 3-6, suggesting that the risk of AMD is heavily influenced by genetic components, and twin studies have estimated the heritability of advanced AMD to be as high as 0.71 (3). Common variants in several alternative complement pathway genes, including complement factor H (*CFH*) (4-9), complement component 2 (*C2*) (8, 10), complement factor B (*CFB*) (8, 10), complement component 3 (*C3*) (11), and complement factor I (*CFI*) (12) and a variant in the age-related maculopathy susceptibility 2 (*ARMS2*) gene (13, 14) modulate AMD risk. Genome-wide association studies (GWAS) in large cohorts have also identified common variants in the high-density lipoprotein cholesterol (HDL), extracellular collagen matrix and angiogenesis pathways (15-17). A meta-analysis confirmed the above loci and added new loci for AMD through GWAS and extensive imputation approaches, yielding a total of 19 significant associations in common loci (18).

Despite a rapidly growing list of associations with common variants, there continues to be a large proportion of the heritability of AMD that is unexplained (17, 18), which might be due to undiscovered common variants or rare alleles in the genome. In many instances linking the associated variant to causal risk-conferring functional variation has been challenging (19). Since common variants have survived the effects of purifying negative selection, they often, by necessity, have subtle biochemical or regulatory functions that can be difficult to assess functionally (20). Sequencing can open up the whole spectrum of the allele frequency distribution and detect rare mutations with obvious functional consequences. In fact, recent studies using sequencing approaches have successfully identified several rare functional variants in *CFH*, *C3*, *CFI* and complement component 9 (*C9*) that may have direct impact on the activation of alternative complement cascade (21-25). A cost-effective alternative approach to query the functional variants across the whole exome is to use an exome array, which provides good coverage for functional variants with frequency as low as 0.01%. To examine the spectrum of rare variation in the exome, we genotyped large cohorts of individuals of European ancestry using the Illumina Infinium HumanExome BeadChip with custom content of loci related to AMD (3,214 additional custom variants).

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Variants passing quality control

We genotyped all samples using the Illumina Infinium HumanExome BeadChip with custom content, of which 161,374 (64.2%) variants are polymorphic in our samples and passed quality control. We then categorized 40,087 variants as common (minor allele frequency [MAF] ≥ 1%), and 121,287 variants as rare (MAF < 1%) based on their minor allele frequency in cases and in controls. This platform captured 1.7% (121,287/7,330,859=.017) of possible rare variant sites in the exome as compared to the >7,300,000 rare variant sties identified in release 0.3 of the Exome Aggregation Consortium (ExAC) (26). Among those rare variants, 72,503 variants were nonsynonymous, nonsense or splice-site variants.

Rare variant analysis

We first tested for association with the rare variants included on the Illumina Infinium HumanExome BeadChip with custom content. After strict quality control and filtering, a total of 57,101 variants were tested for association with AMD. A low frequency, nonsynonymous variant in the pellino E3 ubiquitin protein ligase family member 3 gene (*PELI3*), rs145732233 (*PELI3* A307V), was significantly associated at the genome-wide level with AMD (odds ratio [OR] = 0.14, $P = 4.3 \times 10^{-10}$) (**Table 1**). This variant was predicted to be 'possibly damaging' by PolyPhen2. In our samples, rs145732233 has a minor allele frequency of 0.21% in cases versus 0.77% in controls (data not shown). Comparatively, the MAF is 0.54% in the ExAC database, which is lower than its frequency in our control group and higher than its frequency in our case group. This indicates that the protective effect of A307V in the *PELI3* gene is likely to be true, although the effect size might be smaller than the value estimated in our samples. Another rare, nonsynonymous SNP in *CFH*, rs35274867 (*CFH* N1050Y), was also significantly associated at a genome-wide level with AMD (OR = 0.76, $P = 6.2\text{x}10^{-12}$). We tested this SNP for independence of the

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

other known common and rare SNPs in the *CFH* gene and found the association signal did decrease slightly in the conditional analysis, but the association signal was still highly significant ($P = 1.6\text{x}10^{-11}$). We also detected several other loci with suggestive evidence of association ($1\text{x}10^{-05} > P > 5\text{x}10^{-08}$) with AMD (**Supplementary Table S2**).

In addition to testing each single-variant individually, we also performed gene-based tests to further investigate the cumulative effects of rare functional variants in AMD. Even though the distribution of association statistics of gene-based tests were slightly deflated, possibly due to lack of power from genes with small numbers of rare variants, we were still able to detect significant signals in several genes (inflation factor $\lambda_{gc} = 0.96$, **Supplementary Figure S1**). We identified a burden of rare variants in the *PELI3* gene ($P = 4.3\text{x}10^{-07}$) using the simple burden analysis. We also found cumulative effects of rare variants in five other genes to be significantly associated at the genome-wide level with AMD: *CFH*, *C3*, *C9*, abnormal spindle microtubule assembly (*ASPM*), and mutS homolog 5 (*MSH5*) using either the simple burden test or the SKAT analysis (**Table 2**). To test if the significant signals were independent of known variants, we conditioned on nearby known common and rare variants and demonstrated that the signals were mostly driven by the known variants.

We additionally examined the association signals at previously reported rare AMD loci in *CFH*, *CFI*, *C3* and *C9* in our recent targeted sequencing study, from which there are a subset of samples that overlap with those used for this genotyping study, and found the results supported those obtained from the prior analysis (21, 22) (**Supplementary Table S1**).

Common variant analysis

We evaluated common variants using logistic regression, adjusting for genetic ancestry based on principal component analysis, analyzing the samples from the Boston, French and Finnish cohorts separately and then performed a meta-analysis to assess the pooled effect of these variants across the three countries. We

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plotted the *P*-values of 15,671 ancestry-informative markers in Quantile-Quantile plots and observed no statistical inflation in the distribution of the association statistic for any of the three country-of-origin specific logistic regression analyses (genomic inflation factor $\lambda_{gc} = 1.0$).

Table 3 shows the newly associated, common variants identified in the advanced AMD analysis. We identified one independent significant signal associated with advanced AMD on chromosome 16 near the chymotrypsinogen B1 gene (*CTRB1)* (logistic regression: OR = 0.71, $P = 1.8\text{x}10^{-07}$; Firth corrected OR = 0.64, $P = 9.6\text{x}10^{-11}$; **Figure 1**) which is not close to any of the known loci. The top variant in this region, rs8056814, is an intergenic single nucleotide polymorphism upstream from *CTBR1*.

We also identified three additional suggestive common variant associations with advanced AMD (**Table 3**). These variants included: a missense variant, rs1801689 in apolipoprotein H (*APOH*) (C325R); an intronic variant, rs11884770 in collagen, type IV, alpha 3 (*COL4A3*); and a synonymous variant, rs4072037 in mucin 1, cell surface associated (*MUC1*). We identified seven variants with a borderline suggestive association with AMD, including variants in genes involved in the complement and the innate immune system pathways, two pathways known to be involved in the etiology of AMD (**Supplementary Table S2**). **Supplementary Table S3** shows common variants or their proxy at 21 loci in the inflammatory/immune, angiogenesis, collagen/extracellular matrix, and lipid pathways previously identified by several GWAS analyses (4, 15-18). All known alleles showed similar effect size and direction in this study as compared to previously published values.

For the AMD subtype analyses (**Supplementary Table S4**), we found no new genome-wide significant associations between the two advanced subtypes, geographic atrophy and neovascular disease, nor did we find any new genome-wide significant association between either of the subtypes and the control group. We did, however, identify a genome-wide suggestive association between geographic atrophy and the

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

missense variant, rs1715828 in dynein assembly factor with WDR repeat domains 1 (*DAW1*) (T121S) ($P = 5.5x10^{-06}$). Genome-wide suggestive associations with neovascular disease include: rs7604613, an intragenic variant near tetratricopeptide repeat domain 32 (*TTC32*) ($P = 1.7x10^{-05}$) and both rs11884770 ($P = 1.7x10^{-05}$) and rs1801689 ($P = 1.9x10^{-06}$), the same variants found to be suggestively associated with advanced AMD. The missense variant rs1801689 in *APOH* shows a stronger association with neovascular disease than with advanced AMD ($OR_{neovascular} = 0.61$ vs. $OR_{overall_advanced} = 0.67$), while rs11884770 in *COL4A3* shows a weaker association with neovascular disease than with overall combined types of advanced AMD ($OR_{neovascular} = 0.81$ vs. $OR_{overall_advanced} = 0.83$). Confirming known genome-wide associations with AMD, we identified genome-wide significant associations between geographic atrophy and the known loci including *CFH*, *C3*, *C2/CFB*, and *ARMS2* and between neovascular disease and known AMD loci including *CFH*, *COL8A1*, *C2/CFB*, *C9*, *TGFBR1*, *CETP*, *C3*, and *TIMP3*. The association analysis testing the difference between the two subtypes identified and confirmed the known association with *ARMS2* as previously reported (27, 28), but did not identify any novel associations with any other variants.

DISCUSSION

In this study, we aimed to find new genetic factors for advanced AMD by querying common and rare functional variants across the exome in a large number of subjects from three cohorts of European ancestry. We identified significant associations between AMD and a rare or low frequency protective missense variants, *PELI3* A307V ($P = 4.3\times10^{-10}$) and *CFH* N1050Y ($P = 6.2x10^{-12}$) and a suggestively associated common protective variant near *CTRB1* (rs8056814, $P = 1.8x10^{-07}$).

The rare non-synonymous variant, rs145732233 in *PELI3* is newly identified in this study and is associated with a decrease in risk of AMD. A burden of rare variants in *PELI3* was also detected in the simple burden test, but just missed the cutoff for significance in the sequence kernel association test (SKAT) analysis. The simple burden test also shows the overall association signal from *PELI3* is

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protective (OR = 0.28, data not shown). *PELI3* encodes E3 ubiquitin ligase pellino, a scaffold protein that helps transmit the immune response signals. Pellino E3 has been demonstrated to augment the expression of type I interferon but not of proinflammatory cytokines in response to toll-like receptor 3 protein (TLR3) activation (29). Similar to the complement pathway, the Toll-like receptors (TLRs) participate in protective response against microbial invasion when activated normally, but could be harmful to the host when activated improperly or uncontrolled (30). A protective variant such as *PELI3* A307V might enhance the signaling transduction between the TLR3 and IRF7 pathways, resulting in downregulation of type I interferon expression (29). Thus, individuals with the *PELI3* A307V mutation may have less severe type I interferon response than individuals with normal pellino E3. Individuals with the *PELI3* A307V mutation may be protected from damage caused by immune response activated abnormally, and therefore could be less likely to develop advanced AMD. Although variants in *TLR3* (31) and toll-like receptor 3 (*TLR4*) (32) have been suspected to be related to AMD, these common associations were not replicated in studies with large cohorts (15-18). By genotyping a large number of samples using the exome array, this study identified a novel rare mutation in a gene that intermediates signals between Toll-like receptors and the innate immune pathway. Further studies of functional roles of this mutation and possible mechanisms associated with AMD are warranted, but considering the results from the rare variant analyses, variants in *PELI3* may play a protective role in AMD.

Rare variant analysis in this study identified a second protective rare variant significantly associated with a decrease in risk of AMD. The missense SNP, rs35274867 (*CFH* N1050Y) (OR = 0.76; $P = 6.2\text{x}10^{-12}$) is located in *CFH*, a gene in the complement pathway with strong associations with risk of AMD. An association analysis of the *CFH* N1060Y variant conditional on the associated risk variants in *CFH* shows this SNP has an independent effect on AMD (**Table 1**). Haplotype construction using the Boston cohort also shows the protective allele at *CFH* N1050Y occurs independently of the risk variants from *CFH* R1210C (**Table 4**). Other studies have reported an association between *CFH* variants, both common and rare, and AMD (4-9, 22, 33, 34). This is the first time *CFH* N1050Y has been significantly and

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

independently associated at the genome-wide level with AMD. This variant was identified in our previous sequencing studies, but was not independently associated at the genome-wide level with AMD at the time, though the variant was noted as having a higher frequency in controls than in AMD cases (22, 25). Additionally, the variant has previously been found with a higher frequency in controls compared to cases in a study of systemic lupus erythematosus (35), and was found at a higher frequency in controls than cases in a study of atypical hemolytic uremic syndrome, a disease that shares genetic influences with AMD. Conversely, there is evidence of affected patients carrying the variants including a father and daughter with basal laminar drusen (36), though this study did not look at the frequency of the variant in controls, nor did it attribute the variant to causing the disease in these two patients, and two patients with hemolytic uremic syndrome (37). The variant is located in complement control protein module 18 (CCP18), a module that harbors multiple disease-associated loci and may to be part of an essential binding site (38). Alteration of binding domains of *CFH* may modify how the factor H protein regulates the alternative complement pathway. As this variant is associated with a decrease in risk of AMD, it suggests the mutation may create a comparatively enhanced binding site. Further study is needed to determine the underlying biological mechanism and protective function of this variant.

The burden analyses detected a suggestive association signal ($P_{\text{Burden}} = 3.2\text{x}10^{-05}$, $P_{\text{SKAT}} = 1.0\text{x}10^{-02}$) in the *CFI* gene, but not as strong as the association signal we recently found in our sequencing study by simple burden tests (21). This may be due to the fact that there were only 12 *CFI* rare variants detected on the exome array, much less compared to the 59 *CFI* rare variants we detected by targeted sequencing. This exemplifies the value of employing a variety of genotyping and sequencing platforms for gene discovery.

We found an association between AMD and a common protective variant rs8056814 in *CTRB1*. This SNP is about 330kb away from rs8053796 in contactin associated protein-like 4 (*CNTNAP4*), where we detected a suggestive association signal ($P = 1.7\text{x}10^{-05}$) in our previous meta-GWAS study (17). Previous studies have reported nominal associations between rs8056814 and AMD in the same direction of effect,

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and a recent, independent but concurrent study, reported a genome-wide significant association between rs8056814 and AMD ($P = 2.9x10^{-10}$), also in the same direction of effect (18, 39-41). In our study, the association signal in the *CTRB1* locus reached statistical significance ($P < 1.24x10^{-06}$) for the first time. Compared with previous studies, rs8056814 was directly genotyped in our analysis instead of being imputed, which enabled us to assess the association at this locus more accurately and with more power. Therefore, the effect of this locus might be underestimated in studies using imputed datasets as the linkage disequilibrium structure may not perfectly capture the genotype information of this locus.

The rs8056814 variant is located 557bp upstream of *CTRB1*, encoding chymotrypsinogen, a serine protease that is secreted into the gastrointestinal tract and activated by proteolytic cleavage with trypsin. *CTRB1* is a known risk locus for type 1 diabetes (42) and it is reported to be associated with variation in HDL levels as well (43). Querying the GTEx database, we found that rs8056814 is a cis-expression quantitative trait locus (eQTL) for the downstream gene, *BCAR1* in whole blood ($P = 1.5x10^{-07}$, $\beta = 0.37$) (The data was obtained from the GTEx Portal and dbGaP accession number phs000424.v6.p1) (44). CHIPseq results from the ENCODE project suggests rs8056814 lies in the promoter region of *CTRB1* in a hepatic carcinoma cell line and in an enhancer region in blood cell lines which could explain the eQTL association with *BCAR1* in blood (45). Eye tissue was not assessed as part of the GTEx and ENCODE projects, so further work is required to determine the role this variant plays in ocular tissues. Further evidence shows another variant, rs7202877, is in high LD with rs8056814 ($r^2 = 0.9$) and regulates expression levels of *CTRB1* (46). If *CTRB1* does play a role in the HDL pathway as suggested by Dastani et al. (43), the role of HDL pathway genes in AMD is not unprecedented. Variants in several HDL pathway genes, such as hepatic lipase (*LIPC*) (15, 16), plasma cholesteryl ester transfer protein (*CETP*) (15-17) and ATP binding cassette subfamily A member 1 (*ABCA1*) (15, 16, 47, 48) show significant associations with AMD. The apolipoprotein E (*APOE*) gene, involved in a different part of the lipid pathway, is also related to AMD (49). Of note, the minor allele for the associated variant in *LIPC* (rs10468017), which is associated with higher HDL levels, also shows a protective effect in AMD (15,

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

16). Considering the effects of other variants in this region, it is possible that the rare allele of rs8056814 could influence metabolic changes in serum glucose and HDL levels through the modulation of *CTRB1* expression levels.

The Illumina Infinium HumanExome BeadChip provides an alternative to whole exome sequencing as a way to assess functional variants across the exome in an unbiased manner. However, it does not provide complete coverage of all functional variants in all genes. The ultra-rare variants (MAF < 0.03%) are not covered by this array, thus limiting our power to detect individual or cumulative effects of these variants. For example, it includes only 12 rare variants in the *CFI* gene, while our recent targeted sequencing analysis detected 59 rare variants (21). The lack of information on those ultra-rare variants could weaken the association signals in the gene-based test, especially in the scenario when most of those ultra-rare variants in this gene are causal and associated with AMD in the same direction.

In summary, we have identified protective AMD loci in *CFH*, *PELI3* and near *CTRB1*. We provide new suggestive loci worthy of follow-up, including common variants in *COL4A3*, *APOH*, and *MUC1*, as well as several rare variants. We also confirmed previously published common loci in several pathways identified by GWAS, and the recently associated rare AMD loci in the complement genes *CFH*, *CFI*, *C3*, and *C9* discovered by targeted sequencing. The new genetic loci associated at the genome-wide level with AMD suggest that genes in other branches of the innate immune pathway and additional genes in the HDL pathway may also be involved in the etiology of AMD. As the knowledge about the genetic architecture of AMD expands, new variants may enhance predictive models (50, 51), and could lead to the development of new therapeutic targets.

MATERIALS AND METHODS

Case-control definitions

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

All individuals were evaluated by a board-certified ophthalmologist who conducted ocular examinations including with visual acuity measurements, dilated slit-lamp biomicroscopy, and stereoscopic color fundus photography. Ophthalmologic medical records and ocular images were also reviewed. All subjects were graded using the Clinical Age-Related Maculopathy Staging (CARMS) system (52). Case patients had either geographic atrophy (advanced central or non-central non-exudative AMD or CARMS grade 4) or neovascular disease (neovascular AMD or CARMS grade 5). Controls did not have early, intermediate or advanced macular degeneration, and were categorized as CARMS grade 1. All controls were ≥ 60 years old.

Boston cohorts were recruited at the Tufts Medical Center in Boston, Massachusetts, U.S.A., and throughout the country through ongoing AMD study protocols, as previously described (3, 8, 11, 12, 15, 53-55). We selected 3,772 unrelated individuals (2,488 case and 1,284 controls) from our large collection of advanced AMD case-control and family cohorts. The French cohort of 1,544 cases and 289 controls was recruited at Hôpital Intercommunal de Créteil, Créteil, France, as previously described (15, 17). The Finnish cohort of 300 cases and 160 controls was recruited at the Helsinki University Central Hospital, Helsinki, Finland and >20,000 additional Finnish controls were recruited as part of The National FINRISK Study, a cross-sectional survey of the Finnish population to assess chronic disease in people aged 25 to 74 (56, 57). We included genotype data for shared controls of 2,909 samples that had been genotyped at the Broad Institute (29, 58-61) (**Supplementary Table S5**). Individuals from all cohorts were self-reported white individuals of European descent. We used the first five principal components generated by EIGENSTRAT (62) based on the ancestry informative markers to calculate Euclidean distances between samples in the Boston and French cohorts and shared control samples. We then randomly selected individual case samples in these cohorts and assigned the nearest unassigned shared controls to the selected case's cohort. We matched 2,434 of these shared controls to the Boston cohort and 475 of these shared controls to the French cohort. The two cohorts with matched controls were re-

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

examined with five outlier removal iterations in EIGENSTRAT to ensure that samples were matched properly by their ethnic background.

Whole exome array genotyping of coding variants

Genotyping was performed using the Illumina Infinium HumanExome BeadChip (v1.0), which provides coverage of over 240,000 functional exonic variants selected from >12,000 whole exome and known variants associated with complex traits in previous GWAS, human leukocyte antigen tags, ancestry-informative markers, markers for identity-by-descent estimation and random synonymous single nucleotide polymorphisms (SNPs) (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). In addition, we customized our assay by adding 3,214 SNPs from candidate AMD genes and genes in associated pathways. Included in the custom content are common variants which achieved a *P*-value less than 0.001 in our previous meta-GWAS studies (15, 17) and 20 common SNPs reported by the AMDGENE consortium meta-GWAS study (18). We conducted genotyping of the Boston, French and Finnish cohort samples at the John Hopkins Genotyping Core Laboratory. We genotyped shared control samples separately at the Broad Institute using the same genotyping platform and custom content as was used for the Boston, French, and Finnish cohort samples. We called genotypes using Illumina's GenomeStudio software and then used zCall (63), a rare-variant caller developed at the Broad Institute, to recover missed rare genotypes.

Statistical Analyses

We required that samples have <2% missing genotype calls for common variants (MAF > 5%) before applying zCall. Then after applying zCall we removed duplicate variants, monomorphic variants, variants with a low call rate (<98%), and variants failing Hardy-Weinberg ($P < 10^{-06}$). We merged genotype calls from the different cohorts by only including variants that passed quality control and passed the Hardy-Weinberg test ($P \geq 10^{-06}$) across all samples. To eliminate any batch effect, we excluded 412 variants with

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

allele frequencies significantly different between the examined controls genotyped at the John Hopkins Core Laboratory and the shared controls genotyped at the Broad Institute ($P < 10^{-03}$). We identified 15,671 ancestry informative markers with high minor allele frequencies (MAF > 5%), and excluded regions near (<1Mb) any of the 20 known AMD loci (18) and the major histocompatibility complex locus (chr 6, 25.0-35.0 Mb). We then pruned the resulting set of variants using the --indep option in PLINK with default parameters (variance inflation factor = 2, window size = 50 SNPs) (64). We assessed relatedness by calculating genome-wide proportion identity-by-descent estimates (PIHAT values) using these ancestry informative markers (64). We identified pairs of sequenced individuals with PIHAT > 0.2, and removed one of those individuals from the analysis. We then used EIGENSTRAT (62) to generate the first 10 principal components based on the ancestry informative markers. We only included shared controls who matched the genetic background of Boston and French samples based on principal components as described previously (21).

For statistical analysis of common variants, we tested for associations assuming an additive genetic model using logistic regression adjusting for the first 10 principle components from the EIGENSTRAT analysis. The summary data were then meta-analyzed using METAL (65). Firth logistic regression analysis, adjusted for the top 10 principal components was performed on the most significant associations using the logistf package in R. In addition to testing each variant for an association with advanced AMD, we also conducted subtype analyses and tested each variant for an association with the two advanced forms of AMD: neovascular disease and geographic atrophy. To recognize independent association signals, we also performed conditional analysis for the significant variants within 1Mb of any of the 20 known AMD loci (**Supplementary Table S3**) by adjusting for the genotype of the adjacent known variant. Genome-wide significance (P-value $< 5\times10^{-08}$) and genome-wide suggestive (P-value $< 1 \times10^{-05}$) thresholds were used evaluate the association signals of common variants.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For rare functional variants we carried out single-variant association tests using the same statistical framework of exact statistics described previously (21, 22). Briefly, we used a 2x2 Fisher's exact test to calculate a one-tailed exact *P*-value for multiple case-control cohorts. We performed analyses on 57,101 variants with a MAF <1%, and included. The first 10 principal components of EIGENSTRAT were included as covariates in this analysis. We applied the same statistical framework on data further stratified by genotypes of nearby known common variants, in addition to country of sample collection, to achieve the *P*-values of conditional analysis for the rare variants. To eliminate potential false positives due to low quality calls of rare variants, we also re-examined the cluster plots of genotype calls for the significant variants after association tests, and excluded variants poorly clustered. Genome-wide significance (*P*-value $< 5\times10^{-08}$) and genome-wide suggestive (*P*-value $< 1 \times10^{-05}$) thresholds were used evaluate the association signals of common variants.

For gene-based analysis, we performed a SKAT analysis on each cohort separately followed by a meta-analysis implemented using the RAREMETAL software, to assess the pooled effect of the rare variant burden across the three cohorts (66). SKAT has been shown to perform well in scenarios when a large fraction of the variants in a region are non-causal or the effects of causal variants are in different directions. We performed analyses using default weights (67) on 72,503 nonsynonymous, nonsense or splice-site variants with MAF between 1 and 0.03% in cases or in controls groups. We excluded extremely rare variants (minor allele count < 5, or MAF < 0.03%) from this test. These variants are located in 12,122 genes. Each gene contains at least two variants passing quality control. The first 10 principal components of EIGENSTRAT were included as covariates in the SKAT analysis. We also carried out conditional analyses by including the minor allele count of nearby known common or rare variants as covariates. Additionally we performed a simple burden test to assess if rare variants were enriched in cases versus controls or in controls versus cases. We used the Fisher's exact statistical framework described for the single-variant association analyses above. To interpret statistical

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

significance, we applied a Bonferroni corrected significance threshold of $P < 2.06\times10^{-6}$ ($P = 0.05/12{,}122$ x2 gene-based tests).

Haplotype construction of and frequency calculation of the *CFH* region was performed using the samples in the Boston cohort. Haplotypes were constructed using the --hap-freq function in PLINK and haplotypes with a frequency of greater than 0.001 in either cases or controls were reported (64). Seven SNPs in *CFH* were used in the haplotype construction including the rare *CFH* SNPs, R1210C and N1050Y.

We used the CaTS-Power Calculator software (http://csg.sph.umich.edu//abecasis/cats/calculator.html) to estimate the power to detect each of the rare single-variant associations identified, assuming a significance threshold of P-value $< 5\times10^{-08}$.

FUNDING

This work was supported in part by the National Institutes of Health [RO1-EY11309, 1R01AR063759, U19 AI111224-01], Bethesda, MD; Massachusetts Lions Eye Research Fund, Inc.; Unrestricted grants from Research to Prevent Blindness, Inc., New York, NY; Doris Duke Charitable Foundation Grant [#2013097]; and the Macular Degeneration Research Fund of the Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Tufts University School of Medicine, Boston, MA.

ACKNOWLEDGEMENTS

We thank the numerous ophthalmologists throughout the United States that contributed to the Boston AMD cohort, the ophthalmologists from the Clinical Researches Functional Unit, CHI Créteil, France, and the ophthalmologists from the Helsinki University Central Hospital, Helsinki, Finland.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CONFLICT OF INTEREST STATEMENT

None.

REFERENCES

1 Sobrin, L. and Seddon, J.M. (2014) Nature and nurture- genes and environment- predict onset and progression of macular degeneration. *Prog. Retin. Eye Res.*, **40**, 1-15.

2 Friedman, D.S., O'Colmain, B.J., Munoz, B., Tomany, S.C., McCarty, C., de Jong, P.T., Nemesure, B., Mitchell, P., Kempen, J. and Eye Diseases Prevalence Research, G. (2004) Prevalence of age-related macular degeneration in the United States. *Arch. Ophthalmol.*, **122**, 564-572.

3 Seddon, J.M., Cote, J., Page, W.F., Aggen, S.H. and Neale, M.C. (2005) The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch. Ophthalmol.*, **123**, 321-327.

4 Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T. *et al.* (2005) Complement factor H polymorphism in age-related macular degeneration. *Science*, **308**, 385-389.

5 Hageman, G.S., Anderson, D.H., Johnson, L.V., Hancox, L.S., Taiber, A.J., Hardisty, L.I., Hageman, J.L., Stockman, H.A., Borchardt, J.D., Gehrs, K.M. *et al.* (2005) A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 7227-7232.

6 Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P., Spencer, K.L., Kwan, S.Y., Noureddine, M., Gilbert, J.R. *et al.* (2005) Complement factor H variant increases the risk of age-related macular degeneration. *Science*, **308**, 419-421.

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

8 Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M.J. and Seddon, J.M. (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat. Genet.*, **38**, 1055-1059.

9 Li, M., Atmaca-Sonmez, P., Othman, M., Branham, K.E., Khanna, R., Wade, M.S., Li, Y., Liang, L., Zareparsi, S., Swaroop, A. *et al.* (2006) CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat. Genet.*, **38**, 1049-1054.

10 Gold, B., Merriam, J.E., Zernant, J., Hancox, L.S., Taiber, A.J., Gehrs, K., Cramer, K., Neel, J., Bergeron, J., Barile, G.R. *et al.* (2006) Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat. Genet.*, **38**, 458-462.

11 Maller, J.B., Fagerness, J.A., Reynolds, R.C., Neale, B.M., Daly, M.J. and Seddon, J.M. (2007) Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat. Genet.*, **39**, 1200-1201.

12 Fagerness, J.A., Maller, J.B., Neale, B.M., Reynolds, R.C., Daly, M.J. and Seddon, J.M. (2009) Variation near complement factor I is associated with risk of advanced AMD. *Eur. J. Hum. Genet.*, **17**, 100-104.

13 Rivera, A., Fisher, S.A., Fritsche, L.G., Keilhauer, C.N., Lichtner, P., Meitinger, T. and Weber, B.H. (2005) Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum. Mol. Genet.*, **14**, 3227-3236.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

14 Jakobsdottir, J., Conley, Y.P., Weeks, D.E., Mah, T.S., Ferrell, R.E. and Gorin, M.B. (2005) Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am. J. Hum. Genet.*, **77**, 389-407.
15 Neale, B.M., Fagerness, J., Reynolds, R., Sobrin, L., Parker, M., Raychaudhuri, S., Tan, P.L., Oh, E.C., Merriam, J.E., Souied, E. *et al.* (2010) Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 7395-7400.
16 Chen, W., Stambolian, D., Edwards, A.O., Branham, K.E., Othman, M., Jakobsdottir, J., Tosakulwong, N., Pericak-Vance, M.A., Campochiaro, P.A., Klein, M.L. *et al.* (2010) Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 7401-7406.
17 Yu, Y., Bhangale, T.R., Fagerness, J., Ripke, S., Thorleifsson, G., Tan, P.L., Souied, E.H., Richardson, A.J., Merriam, J.E., Buitendijk, G.H. *et al.* (2011) Common variants near FRK/COL10A1 and VEGFA are associated with advanced age-related macular degeneration. *Hum. Mol. Genet.*, **20**, 3699-3709.
18 Fritsche, L.G., Chen, W., Schu, M., Yaspan, B.L., Yu, Y., Thorleifsson, G., Zack, D.J., Arakawa, S., Cipriani, V., Ripke, S. *et al.* (2013) Seven new loci associated with age-related macular degeneration. *Nat. Genet.*, **45**, 433-439, 439e431-432.
19 Ioannidis, J.P., Thomas, G. and Daly, M.J. (2009) Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.*, **10**, 318-329.
20 Raychaudhuri, S. (2011) Mapping rare and common causal alleles for complex human diseases. *Cell*, **147**, 57-69.
21 Seddon, J.M., Yu, Y., Miller, E.C., Reynolds, R., Tan, P.L., Gowrisankar, S., Goldstein, J.I., Triebwasser, M., Anderson, H.E., Zerbib, J. *et al.* (2013) Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat. Genet.*, **45**, 1366-1370.
22 Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P.L., Tai, A.K., Ripke, S., Gowrisankar, S., Vemuri, S., Montgomery, K., Yu, Y. *et al.* (2011) A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat. Genet.*, **43**, 1232-1236.
23 van de Ven, J.P., Nilsson, S.C., Tan, P.L., Buitendijk, G.H., Ristau, T., Mohlin, F.C., Nabuurs, S.B., Schoenmaker-Koller, F.E., Smailhodzic, D., Campochiaro, P.A. *et al.* (2013) A functional variant in the CFI gene confers a high risk of age-related macular degeneration. *Nat. Genet.*, **45**, 813-817.
24 Kavanagh, D., Yu, Y., Schramm, E.C., Triebwasser, M., Wagner, E.K., Raychaudhuri, S., Daly, M.J., Atkinson, J.P. and Seddon, J.M. (2015) Rare genetic variants in the CFI gene are associated with advanced age-related macular degeneration and commonly result in reduced serum factor I levels. *Hum. Mol. Genet.*, **24**, 3861-3870.
25 Triebwasser, M.P., Roberson, E.D., Yu, Y., Schramm, E.C., Wagner, E.K., Raychaudhuri, S., Seddon, J.M. and Atkinson, J.P. (2015) Rare variants in the functional domains of complement factor H are associated with age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.*, **56**, 6873-6878.
26 Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B. *et al.* (2015) Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*, in press.
27 Sobrin, L., Reynolds, R., Yu, Y., Fagerness, J., Leveziel, N., Bernstein, P.S., Souied, E.H., Daly, M.J. and Seddon, J.M. (2011) ARMS2/HTRA1 locus can confer differential susceptibility to the advanced subtypes of age-related macular degeneration. *Am. J. Ophthalmol.*, **151**, 345-352 e343.
28 Sobrin, L., Ripke, S., Yu, Y., Fagerness, J., Bhangale, T.R., Tan, P.L., Souied, E.H., Buitendijk, G.H., Merriam, J.E., Richardson, A.J. *et al.* (2012) Heritability and genome-wide association study to assess genetic differences between advanced age-related macular degeneration subtypes. *Ophthalmology*, **119**, 1874-1885.
29 Siednienko, J., Jackson, R., Mellett, M., Delagic, N., Yang, S., Wang, B., Tang, L.S., Callanan, J.J., Mahon, B.P. and Moynagh, P.N. (2012) Pellino3 targets the IRF7 pathway and facilitates

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

autoregulation of TLR3- and viral-induced expression of type I interferons. *Nat. Immunol.*, **13**, 1055-1062.
30 Di Domizio, J. and Cao, W. (2013) Fueling autoimmunity: type I interferon in autoimmune diseases. *Expert Rev. Clin. Immunol.*, **9**, 201-210.
31 Yang, Z., Stratton, C., Francis, P.J., Kleinman, M.E., Tan, P.L., Gibbs, D., Tong, Z., Chen, H., Constantine, R., Yang, X. *et al.* (2008) Toll-like receptor 3 and geographic atrophy in age-related macular degeneration. *N. Engl. J. Med.*, **359**, 1456-1463.
32 Zareparsi, S., Buraczynska, M., Branham, K.E., Shah, S., Eng, D., Li, M., Pawar, H., Yashar, B.M., Moroi, S.E., Lichter, P.R. *et al.* (2005) Toll-like receptor 4 variant D299G is associated with susceptibility to age-related macular degeneration. *Hum. Mol. Genet.*, **14**, 1449-1455.
33 Hoffman, J.D., Cooke Bailey, J.N., D'Aoust, L., Cade, W., Ayala-Haedo, J., Fuzzell, D., Laux, R., Adams, L.D., Reinhart-Mercer, L., Caywood, L. *et al.* (2014) Rare complement factor H variant associated with age-related macular degeneration in the Amish. *Invest. Ophthalmol. Vis. Sci.*, **55**, 4455-4460.
34 Yu, Y., Triebwasser, M.P., Wong, E.K., Schramm, E.C., Thomas, B., Reynolds, R., Mardis, E.R., Atkinson, J.P., Daly, M., Raychaudhuri, S. *et al.* (2014) Whole-exome sequencing identifies rare, functional CFH variants in families with macular degeneration. *Hum. Mol. Genet.*, **23**, 5283-5293.
35 Jonsen, A., Nilsson, S.C., Ahlqvist, E., Svenungsson, E., Gunnarsson, I., Eriksson, K.G., Bengtsson, A., Zickert, A., Eloranta, M.L., Truedsson, L. *et al.* (2011) Mutations in genes encoding complement inhibitors CD46 and CFH affect the age at nephritis onset in patients with systemic lupus erythematosus. *Arthritis Research & Therapy*, **13**, R206.
36 Boon, C.J., Klevering, B.J., Hoyng, C.B., Zonneveld-Vrieling, M.N., Nabuurs, S.B., Blokland, E., Cremers, F.P. and den Hollander, A.I. (2008) Basal laminar drusen caused by compound heterozygous variants in the CFH gene. *Am. J. Hum. Genet.*, **82**, 516-523.
37 Richards, A., Buddles, M.R., Donne, R.L., Kaplan, B.S., Kirk, E., Venning, M.C., Tielemans, C.L., Goodship, J.A. and Goodship, T.H. (2001) Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *Am. J. Hum. Genet.*, **68**, 485-490.
38 Blackmore, T.K., Hellwage, J., Sadlon, T.A., Higgs, N., Zipfel, P.F., Ward, H.M. and Gordon, D.L. (1998) Identification of the second heparin-binding domain in human complement factor H. *J. Immunol.*, **160**, 3342-3348.
39 Edwards, A.O., Fridley, B.L., James, K.M., Sharma, A.K., Cunningham, J.M. and Tosakulwong, N. (2008) Evaluation of clustering and genotype distribution for replication in genome wide association studies: the age-related eye disease study. *PLoS One*, **3**, e3813.
40 Ryu, E., Fridley, B.L., Tosakulwong, N., Bailey, K.R. and Edwards, A.O. (2010) Genome-wide association analyses of genetic, phenotypic, and environmental risks in the age-related eye disease study. *Mol. Vis.*, **16**, 2811-2821.
41 Fritsche, L.G., Igl, W., Bailey, J.N., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon, K.P., Hebbring, S.J., Wen, C., Gorski, M. *et al.* (2016) A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.*, **48**, 134-143.
42 Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C. *et al.* (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.*, **41**, 703-707.
43 Dastani, Z., Pajukanta, P., Marcil, M., Rudzicz, N., Ruel, I., Bailey, S.D., Lee, J.C., Lemire, M., Faith, J., Platko, J. *et al.* (2010) Fine mapping and association studies of a high-density lipoprotein cholesterol linkage region on chromosome 16 in French-Canadian subjects. *Eur. J. Hum. Genet.*, **18**, 342-347.
44 Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580-585.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

45 Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
46 't Hart, L.M., Fritsche, A., Nijpels, G., van Leeuwen, N., Donnelly, L.A., Dekker, J.M., Alssema, M., Fadista, J., Carlotti, F., Gjesing, A.P. *et al.* (2013) The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes*, **62**, 3275-3281.
47 Yu, Y., Reynolds, R., Fagerness, J., Rosner, B., Daly, M.J. and Seddon, J.M. (2011) Association of variants in the LIPC and ABCA1 genes with intermediate and large drusen and advanced age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.*, **52**, 4663-4670.
48 Yu, Y., Reynolds, R., Rosner, B., Daly, M.J. and Seddon, J.M. (2012) Prospective assessment of genetic effects on progression to different stages of age-related macular degeneration using multistate Markov models. *Invest Ophthalmol Vis Sci*, **53**, 1548-1556.
49 Souied, E.H., Benlian, P., Amouyel, P., Feingold, J., Lagarde, J.P., Munnich, A., Kaplan, J., Coscas, G. and Soubrane, G. (1998) The epsilon4 allele of the apolipoprotein E gene as a potential protective factor for exudative age-related macular degeneration. *Am. J. Ophthalmol.*, **125**, 353-359.
50 Seddon, J.M., Reynolds, R., Maller, J., Fagerness, J.A., Daly, M.J. and Rosner, B. (2009) Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. *Invest. Ophthalmol. Vis. Sci.*, **50**, 2044-2053.
51 Seddon, J.M., Silver, R.E., Kwong, M. and Rosner, B. (2015) Risk prediction for progression of macular degeneration: 10 common and rare genetic variants, demographic, environmental, and macular covariates. *Invest. Ophthalmol. Vis. Sci.*, **56**, 2192-2202.
52 Seddon, J.M., Sharma, S. and Adelman, R.A. (2006) Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology*, **113**, 260-266.
53 Seddon, J.M., Cote, J., Davis, N. and Rosner, B. (2003) Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio. *Arch. Ophthalmol.*, **121**, 785-792.
54 Seddon, J.M., Santangelo, S.L., Book, K., Chong, S. and Cote, J. (2003) A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. *Am. J. Hum. Genet.*, **73**, 780-790.
55 Seddon, J.M., Rosner, B., Sperduto, R.D., Yannuzzi, L., Haller, J.A., Blair, N.P. and Willett, W. (2001) Dietary fat and risk for advanced age-related macular degeneration. *Arch. Ophthalmol.*, **119**, 1191-1199.
56 Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Mannisto, S., Salomaa, V., Sundvall, J. and Puska, P. (2015) Forty-year trends in cardiovascular risk factors in Finland. *European Journal of Public Health*, **25**, 539-546.
57 Vartiainen, E., Laatikainen, T., Peltonen, M., Juolevi, A., Mannisto, S., Sundvall, J., Jousilahti, P., Salomaa, V., Valsta, L. and Puska, P. (2010) Thirty-five-year trends in cardiovascular risk factors in Finland. *International Journal of Epidemiology*, **39**, 504-518.
58 Consortium, G.P., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061-1073.
59 Daly, A.K., Donaldson, P.T., Bhatnagar, P., Shen, Y., Pe'er, I., Floratos, A., Daly, M.J., Goldstein, D.B., John, S., Nelson, M.R. *et al.* (2009) HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.*, **41**, 816-819.
60 Sklar, P., Smoller, J.W., Fan, J., Ferreira, M.A., Perlis, R.H., Chambert, K., Nimgaonkar, V.L., McQueen, M.B., Faraone, S.V., Kirby, A. *et al.* (2008) Whole-genome association study of bipolar disorder. *Mol. Psychiatry.*, **13**, 558-569.
61 Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N. *et al.* (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.*, **43**, 1066-1073.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

62 Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904-909.

63 Goldstein, J.I., Crenshaw, A., Carey, J., Grant, G.B., Maguire, J., Fromer, M., O'Dushlaine, C., Moran, J.L., Chambert, K., Stevens, C. *et al.* (2012) zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics*, **28**, 2543-2545.

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

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

66 Feng, S., Liu, D., Zhan, X., Wing, M.K. and Abecasis, G.R. (2014) RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics*, **30**, 2828-2829.

67 Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Team, N.G.E.S.P.-E.L.P., Christiani, D.C., Wurfel, M.M. and Lin, X. (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet*, **91**, 224-237.

FIGURE LEGENDS:

Figure 1. Zoomplots summarizing association results for the *CTRB1* locus.

The regional association plot from analysis of common variants in 4,332 cases and 4,642 controls of European ancestry. Gene location is shown along the bottom of the graph, with observed $-\log_{10}$(*P*-value) along the left Y-axis and recombination rate along the right Y-axis. Each variant is plotted as a circle, filled with color coded according to the extent of linkage disequilibrium (variants with missing LD information are shown in grey) with the index variant, rs8056814 (colored in purple).

Supplementary Figure S1. Quantile-Quantile plot of gene-based tests.

We tested a total of 12,122 genes using the sequence kernel association test (SKAT). We plot the observed *P*-value for each gene as a function of expected *P*-values.

Supplementary Figure S2. Quantile-Quantile plot of common ancestry informative markers.

We tested 15,671 ancestry-informative markers excluding regions near (<1Mb) any of the 19 known AMD loci (18), and the major histocompatability complex locus (chr 6, 25-35 MB). We plot the observed *P*-value for each marker as a function of expected *P*-values.

Supplementary Figure S3. Manhattan Plot of common variants.

We plot the $-\log_{10}$(*P*-value) for each common variant organized by location on each chromosome.

Supplementary Figure S4. Quantile-Quantile plot of rare variants.

We plot the observed *P*-value for each rare variant as a function of expected *P*-values.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure S5. Quantile-Quantile plot of common variants for advanced AMD vs unaffected analysis.

We plot the observed *P*-value in the advanced AMD vs unaffected regression analysis for each common variant as a function of expected *P*-values.

Supplementary Figure S6. Quantile-Quantile plot of common variants for neovascular disease vs unaffected analysis.

We plot the observed *P*-value in the neovascular disease vs unaffected regression analysis for each common variant as a function of expected *P*-values.

Supplementary Figure S7. Quantile-Quantile plot of common variants for geographic atrophy vs unaffected analysis.

We plot the observed *P*-value in the geographic atrophy vs unaffected regression analysis for each common variant as a function of expected *P*-values.

Supplementary Figure S8. Quantile-Quantile plot of common variants for geographic atrophy vs neovascular disease analysis.

We plot the observed *P*-value in the geographic atrophy vs neovascular disease analysis for each common variant as a function of expected *P*-values.

Supplementary Figure S9. Principal component analysis including all cases and control samples.

We plot the eigenvector1 values for each case and control sample as a function of engenvector2 values to show no stratification between cases and controls is present in the dataset. Cases = red +; controls = green x.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. New rare or low frequency age-related macular degeneration associated variants: associations in the single-SNP analysis

GENE	SNP	Coordinates[A]	Amino Acid Change	Function	Minor allele	% MAF in ExAC[B]	Boston Cohort		French Cohort		Finnish Cohort		OR[C]	P^D	Conditional P^E	Conditioned on SNPs
							% MAF cases	% MAF controls	% MAF cases	% MAF controls	% MAF cases	% MAF controls				
CFH	rs35274867	1:196712596	N1050Y	Missense	T	1.945	0.583	1.950	1.231	2.029	0.333	0.852	0.76	6.22×10^{-12}**	1.6×10^{-11}	rs10737680; rs1061147; rs121913059
PELI3	rs145732233	11:66243148	A307V	Missense	T	0.539	0.222	0.835	0.228	0.590	0	1.631	0.14	4.30×10^{-10}**	-	-

[A]GRCh37/hg19. [B]Percent minor allele frequency (MAF) in the Exome Aggregation Consortium (ExAC) samples of European ancestry. [C]Odds ratio for exact test stratified by country of sample collection. [D]*P*-value for exact test stratified by country of sample collection. [E]Conditional *P*-value for the exact test stratified by the genotypes of known common SNPs in nearby regions and the country of sample collection. ** Genome-wide significant.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Gene based analysis for burden of rare variants in age-related macular degeneration

Gene	Chromosome	Number of variants[A]	P_{Burden}[B]	P_{SKAT}[C]	Conditional P_{SKAT}[D]	Conditioned on variants
PELI3	11	3	4.3×10^{-07}*	4.3×10^{-06}	5.5×10^{-01}	rs145732233
MSH5	6	7	2.4×10^{-08}*	1.1×10^{-07}*	1.5×10^{-01}	rs429608; rs3129987
C9	5	13	1.1×10^{-04}	9.8×10^{-08}*	3.0×10^{-01}	rs34882957
CFH	1	12	3.8×10^{-12}*	2.8×10^{-15}*	5.1×10^{-01}	rs1061147; rs10737680; rs121913059; rs35274867
ASPM	1	31	1.1×10^{-05}	3.5×10^{-07}*	7.3×10^{-01}	rs1061147; rs10737680; rs121913059; rs35274867
C3	19	10	1.1×10^{-03}	1.7×10^{-12}*	7.9×10^{-01}	rs2230199; rs147859257

*Genes with $P < 2.06 \times 10^{-06}$. [A]Number of rare variants passing stringent quality control in each gene used for each test. [B]*P*-value of gene-based simple burden analysis. [C]*P*-value of gene-based SKAT analysis. [D]*P*-value of the conditional gene-based SKAT analysis for each gene adjusting for the genotypes of common and rare variants in and near the gene.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. New common age-related macular degeneration associated variants: meta-analysis of age-related macular degeneration cohorts

Gene	SNP	Coordinates[A]	Minor allele	Boston cohort				French cohort				Finnish cohort				Meta-analysis			Firth Corrected Meta-analysis	
				MAF cases	MAF controls	OR	*P*	MAF cases	MAF controls	OR	*P*	MAF cases	MAF controls	OR	*P*	OR[B]	*P*[C]	Direction[D]	OR[E]	*P*[F]
CTRB1	rs8056814	16:75252327	A	0.072	0.104	0.67	$2.5x10^{-07}$	0.080	0.088	0.90	$5.1x10^{-01}$	0.063	0.083	0.75	$9.1x10^{-02}$	0.71	$1.8x10^{-07}$*	---	0.64	$9.6 x10^{-11}$
APOH	rs1801689	17:64210580	C	0.028	0.038	0.71	$1.5x10^{-03}$	0.029	0.048	0.56	$3.6x10^{-04}$	0.007	0.008	0.83	$7.2x10^{-01}$	0.67	$4.1x10^{-06}$*	---	0.74	$7.7 x10^{-04}$
COL4A3	rs11884770	2:228086920	T	0.251	0.280	0.85	$1.4x10^{-03}$	0.248	0.315	0.72	$1.3x10^{-03}$	0.308	0.337	0.87	$1.4x10^{-01}$	0.83	$6.6x10^{-06}$*	---	0.88	$4.0 x10^{-03}$
MUC1	rs4072037	1:155162067	C	0.453	0.476	0.89	$1.6x10^{-03}$	0.444	0.495	0.80	$5.8x10^{-04}$	0.423	0.438	0.94	$4.6x10^{-01}$	0.88	$9.6x10^{-06}$*	---	0.91	$1.9 x10^{-03}$

Common variants (minor allele frequency [MAF] ≥ 1%) in cases and in controls across all three populations with significant (*P*-value < $1.24x10^{-06}$) or suggestive (*P* < $2.49x10^{-05}$) associations with advanced AMD as discovered in the meta-analysis. [A]GRCh37/hg19. [B]Odds ratio from the meta-analysis of logistic regression analysis of common variants. [C]*P*-value from the meta-analysis of logistic regression analysis of common variants. [D]Direction of effect from the three populations in the following order: Boston, French, Finnish. [E]Odds ratio from the meta-analysis of Firth logistic regression analysis of common variants. [F]*P*-value from the meta-analysis of Firth logistic regression analysis of common variants. *Genome-wide suggestive.

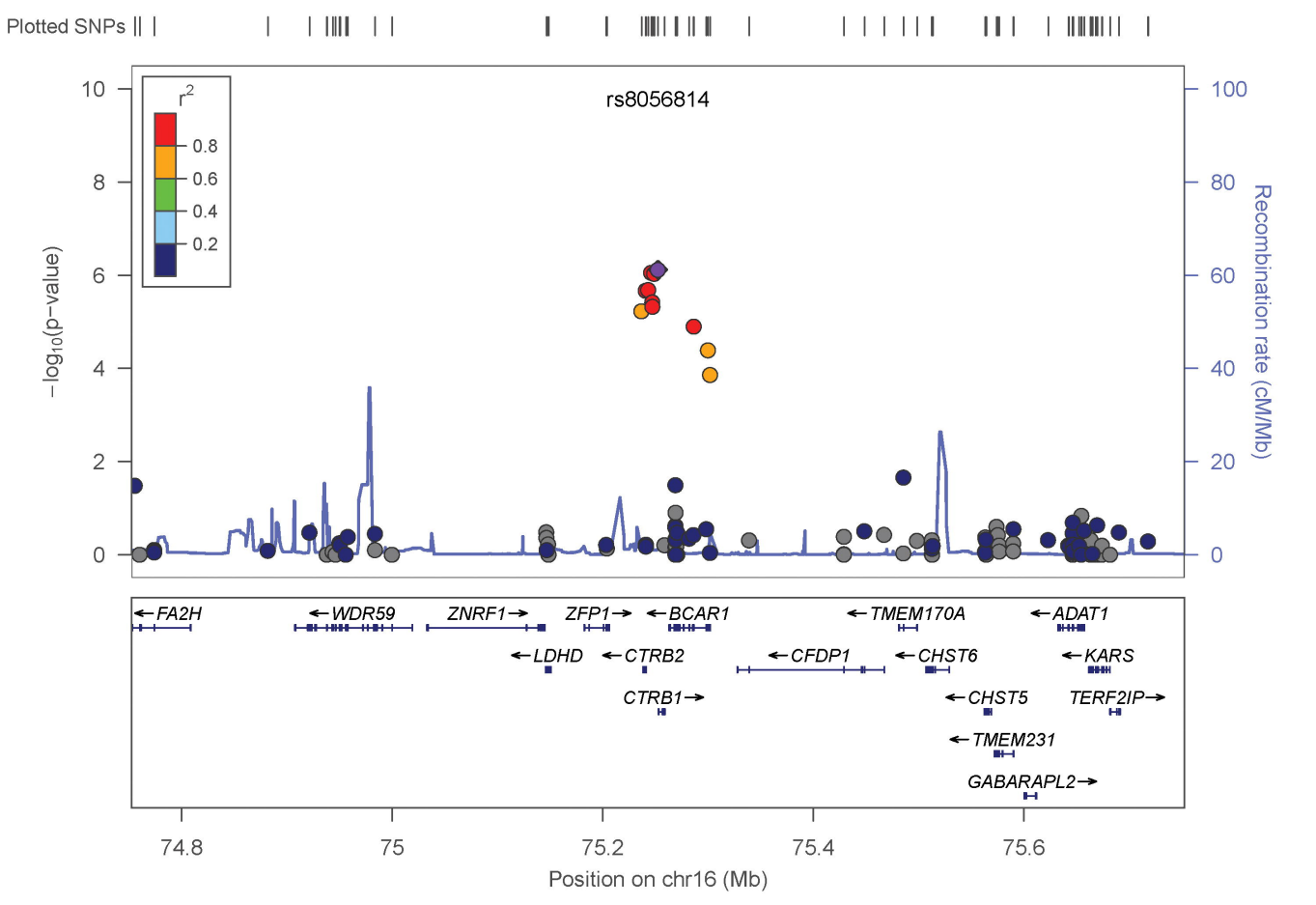
bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 4. *CFH* Haplotype construction in the Boston cohort

	rs800292	rs1329424	rs1061147	rs6428357	rs10737680	rs35274867	rs121913059	Haplotype frequency - cases	Haplotype frequency - controls
Haplotype 1	G	T	A	A	A	A	C	0.5676	0.2815
Haplotype 2	G	G	C	G	A	A	C	0.1464	0.1352
Haplotype 3	A	G	C	G	C	A	C	0.09861	0.1633
Haplotype 4	G	G	C	G	C	A	C	0.08345	0.1493
Haplotype 5	G	G	C	A	A	A	C	0.04589	0.06361
Haplotype 6	G	T	A	G	A	A	C	0.03085	0.06982
Haplotype 7	A	G	C	A	C	A	C	0.002634	0.05047
Haplotype 8	G	G	C	A	C	A	C	0.00345	0.04411
Haplotype 9	G	G	C	G	C	T[A]	C	0.005384	0.01624
Haplotype 10	A	G	C	G	A	A	C	0.004969	0.01337
Haplotype 11	G	G	C	G	C	A	T[B]	0.005383	0.000287
Haplotype 12	G	G	C	A	C	T[A]	C	0.000444	0.003816
Haplotype 13	A	G	C	A	A	A	C	0.001758	0.001601
Haplotype 14	A	G	C	A	A	A	C	0.000552	0.001311
Haplotype 15	A	T	A	G	C	A	C	0.000402	0.002579
Haplotype 16	G	T	A	G	C	A	C	0.000603	0.001719

[A]Protective allele at *CFH* N1050Y. [B]Risk allele at *CFH* R1210C.

bioRxiv preprint doi: https://doi.org/10.1101/034173; this version posted April 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


Plotted SNPs
rs8056814
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FA2H
WDR59
ZNRF1
ZFP1
BCAR1
TMEM170A
ADAT1
LDHD
CTRB2
CFDP1
CHST6
KARS
CTRB1
CHST5
TERF2IP
TMEM231
GABARAPL2
74.8
75
75.2
75.4
75.6
Position on chr16 (Mb)