Human Molecular Genetics, 2016, Vol. 25, No. 23 5276–5285

doi: 10.1093/hmg/ddw336
Advance Access Publication Date: 22 December 2016
Association Studies Article

ASSOCIATION STUDIES ARTICLE

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

Yi Yu[1], Erin K. Wagner[1,2], Eric H. Souied[3], Sanna Seitsonen[4], Ilkka J. Immonen[4], Paavo Häppölä[5], Soumya Raychaudhuri[6,7,8,9,10], Mark J. Daly[6,7,11] and Johanna M. Seddon[1,2,12,*]

[1]Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, MA, USA, [2]Department of Ophthalmology, Tufts University School of Medicine, Boston, MA, USA, [3]Hôpital Intercommunal, Hôpital Henri Mondor, Créteil Université Paris Est, Paris, France, [4]Helsinki University Eye Clinic, Helsinki, Finland, [5]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland, [6]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, [7]Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA, [8]Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA, [9]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA, [10]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK, [11]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA and [12]Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, MA, USA

*To whom correspondence should be addressed at: Johanna M. Seddon, Tufts Medical Center, 800 Washington Street, #450, Boston, MA 02111, USA. Tel: +1 6176369000; Fax: +1 6176365844; Email: jseddon@tuftsmedicalcenter.org

Abstract

Although numerous common age-related macular degeneration (AMD) alleles have been discovered using genome-wide association studies, substantial disease heritability remains unexplained. We sought to identify additional common and rare variants associated with advanced AMD. A total of 4,332 cases and 25,268 controls of European ancestry from three different populations were genotyped using the Illumina Infinium HumanExome BeadChip. We performed meta-analyses to identify associations with common variants, and single variant and gene-based burden tests to identify rare variants. Two protective, low-frequency, non-synonymous variants were significantly associated with a decrease in AMD risk: A307V in *PELI3* (odds ratio [OR] = 0.14, $P = 4.3 \times 10^{-10}$) and N1050Y in *CFH* (OR = 0.76, $P = 6.2 \times 10^{-12}$). The new variants have a large effect size, similar to some rare mutations we reported previously in a targeted sequencing study, which remain significant in this analysis: *CFH* R1210C (OR = 18.82, $P = 3.5 \times 10^{-07}$), *C3* K155Q (OR = 3.27, $P = 1.5 \times 10^{-10}$) and *C9* P167S (OR = 2.04, $P = 2.8 \times 10^{-07}$). 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, $P = 1.8 \times 10^{-07}$). Suggestive protective loci were identified in the *COL4A3* and *APOH* genes. Our results support the involvement of common and low-frequency protective variants in this vision-threatening condition. This study expands the roles of the innate immune pathway as well as the extracellular matrix and high-density lipoprotein pathways in the aetiology of AMD.

†Presented at the Association for Research in Vision and Ophthalmology Meeting, Seattle, Washington, May 2, 2016.
Received: January 26, 2016. **Revised:** September 16, 2016. **Accepted:** September 29, 2016

Introduction

Advanced age-related macular degeneration (AMD) (MIM 603075) is a common, complex and a chronic eye disease (1). 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 by 2020 due to aging of the population (2). As the prevalence of AMD expands, the personal, societal and economic burden of the disease rises. 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), as well as 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 (ECM) 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, a large proportion of the heritability of AMD remains unexplained (17,18), likely due to undiscovered common variants or rare alleles in the genome. In many instances, however, linking associated variants 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 provide accessibility to the entire spectrum of 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).

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 were polymorphic in the Boston and French samples and passed quality control. We then categorized 40,087 variants as common (minor allele frequency [MAF] $\geq$ 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 = 0.017) of possible rare variant sites in the exome as compared to the >7,300,000 rare variant sites identified in release 0.3 of the Exome Aggregation Consortium (ExAC) (26). Among those rare variants, 72,503 variants were non-synonymous, nonsense, or splice-site variants.

Rare variant analysis

We first tested the association of AMD 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 included (Quantile-Quantile [QQ] plot of rare single nucleotide polymorphisms [SNPs] shown in Supplementary Material, Fig. S1). A low-frequency, non-synonymous variant in the pellino E3 ubiquitin protein ligase family member 3 gene (*PELI3*), rs145732233 (*PELI3* A307V), was significantly associated with AMD (odds ratio [OR] = 0.14, $P = 4.3 \times 10^{-10}$; Firth OR = 0.30, $P = 5.9 \times 10^{-07}$) (Table 1). This variant was predicted to be 'possibly damaging' by PolyPhen2. In the ExAC database, the MAF is 0.54%, 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, non-synonymous SNP in *CFH*, rs35274867 (*CFH* N1050Y), was also significantly associated with a reduced risk of AMD (OR = 0.76, $P = 6.2 \times 10^{-12}$; Firth OR = 0.40, $P = 1.6 \times 10^{-11}$). We tested whether this SNP was independent of the other known common and rare SNPs in the *CFH* gene and found the association signal was still highly significant ($P_{conditional} = 1.6 \times 10^{-11}$).

In addition to testing each variant individually, we performed gene-based tests to further investigate the cumulative effects of rare functional variants in AMD. Although 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 (Supplementary Material, Fig. S2). We identified a burden of rare variants in the *PELI3* gene ($P = 4.3 \times 10^{-07}$) using the simple burden analysis which was explained primarily by the newly discovered rare variant, rs145732233 (A307V). Using either the simple burden test or the SKAT analysis, we also found cumulative effects of rare variants in five other genes to be significantly associated with AMD: *CFH*, *C3*, *C9*, abnormal spindle microtubule assembly (*ASPM*), and mutS homolog 5 (*MSH5*). We conditioned on nearby known common and rare variants and demonstrated that the signals found in these five genes were mostly driven by known variants (Table 2).

We additionally examined association signals at the rare AMD loci in *CFH*, *CFI*, *C3* and *C9* reported in our previous studies, and a subset of these samples is included in this larger cohort. Our results supported the associations of these rare variants with AMD (21,22,24) (Supplementary Material, Table S1).

Common variant analysis

We evaluated common variants using logistic regression, adjusting for genetic ancestry based on principal component analysis. We analyzed samples from the Boston, French and Finnish cohorts separately and then performed a meta-analysis to assess the pooled effect of these variants. After excluding the custom SNPs (3,214) and SNPs in regions near (<1Mb) any of the 19 known AMD loci (18) and the major histocompatibility complex locus (chr 6, 25.0–35.0 Mb), we plotted the *P*-values of the remaining 32,958 common variants. We observed no statistical

Table 1. 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	Firth bias corrected analysis	
							% MAF cases	% MAF controls	% MAF cases	% MAF controls	% MAF cases	% MAF controls					OR[f]	P^g
CFH	rs35274867	1:196712596	N1050Y	Missense	T	1.945	0.583	1.950	1.231	2.029	0.333	0.852	0.76	6.2×10^{-12}**	1.6×10^{-11}	rs10737680; rs1061147; rs121913059	0.40	1.6×10^{-11}
PELI3	rs145732233	11:66243148	A307V	Missense	T	0.539	0.222	0.835	0.228	0.590	0	1.631	0.14	4.3×10^{-10}**	–	–	0.30	5.9×10^{-07}

**Significant.
[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.
[f]Odds ratio from the Firth logistic regression analysis of rare variants adjusting for first 10 principle components after excluding the additional Finnish controls (>20,000 from The National FINRISK Study).
[g]P-value from the meta-analysis of Firth logistic regression analysis of rare variants adjusting for first 10 principle components after excluding the additional Finnish controls (>20,000 from The National FINRISK Study).

inflation in the distribution of the association statistic for any of the three country-of-origin specific logistic regression analyses (Supplementary Material, Fig. S3). The Manhattan plot of common variants is shown in (Supplementary Material, Fig. S4.).

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, $P = 1.8\times10^{-07}$; Fig. 1) that is not close to any of the known loci. The top variant in this region, rs8056814, is an intergenic SNP upstream from *CTRB1*.

We identified three additional common variants that showed a suggestive association with reduced risk of advanced AMD (Table 3). These variants included: missense variant rs1801689 (C325R) in apolipoprotein H (*APOH*); 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 (vitronectin [*VTN*], decay accelerating factor for complement [*CD55*] and complement component receptor 2 [*CR2*]), two pathways known to be involved in the aetiology of AMD (Supplementary Material, Table S2). Supplementary Material, Table S3 shows results for 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 (4,8,15–18). All known alleles showed similar effect size and direction in this study as compared to previously published values.

We evaluated the two major subtypes of advanced AMD and found no new significant genetic associations when separately comparing neovascular disease or geographic atrophy to the control group, or when comparing the two subtypes to each other (QQ plots for each subtype analysis shown in Supplementary Materials, Figs. S5–S7). We did, however, identify a suggestive association between geographic atrophy and the missense variant rs1715828 (T121S) in dynein assembly factor with WDR repeat domains 1, *DAW1* ($P = 5.5\times10^{-06}$). Suggestive associations with neovascular disease include rs7604613, an intragenic variant near tetratricopeptide repeat domain 32, *TTC32* ($P = 1.7\times10^{-05}$) and both *COL4A3* rs11884770 ($P = 1.7\times10^{-05}$) and missense variant *APOH* rs1801689 ($P = 1.9\times10^{-06}$) (Supplementary Material, Table S4), the same variants found to be suggestively associated with overall advanced AMD. We identified 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*. We tested the difference between the two AMD subtypes and confirmed the strong association of *ARMS2* with neovascular disease compared with geographic atrophy as previously reported (27,28), but did not identify novel associations with any other variants.

Discussion

In this study, we aimed to identify 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 observed significant associations between AMD and rare or low-frequency protective missense variants *PELI3* A307V ($P = 4.3\times10^{-10}$) and *CFH* N1050Y

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	$4.3x10^{-07*}$	$4.3x10^{-06}$	$5.5\ x10^{-01}$	rs145732233
MSH5	6	7	$2.4x10^{-08*}$	$1.1x10^{-07*}$	$1.5x10^{-01}$	rs429608; rs3129987
C9	5	13	$1.1x10^{-04}$	$9.8x10^{-08*}$	$3.0x10^{-01}$	rs34882957
CFH	1	12	$3.8x10^{-12*}$	$2.8x10^{-15*}$	$5.1x10^{-01}$	rs1061147; rs10737680; rs121913059; rs35274867
ASPM	1	31	$1.1x10^{-05}$	$3.5x10^{-07*}$	$7.3x10^{-01}$	rs1061147; rs10737680; rs121913059; rs35274867
C3	19	10	$1.1x10^{-03}$	$1.7x10^{-12*}$	$7.9x10^{-01}$	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.

Table 3. 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		
				% MAF cases	% MAF controls	OR	P	% MAF cases	% MAF controls	OR	P	% MAF cases	% MAF controls	OR	P	OR[b]	P[c]	Direction[d]
CTRB1	rs8056814	16:75252327	A	7.2	10.4	0.67	$2.5x10^{-07}$	8.0	8.8	0.90	$5.1x10^{-01}$	6.3	8.3	0.75	$9.1x10^{-02}$	0.71	$1.8x10^{-07}$	- - -
APOH	rs1801689	17:64210580	C	2.8	3.8	0.71	$1.5x10^{-03}$	2.9	4.8	0.56	$3.6x10^{-04}$	0.7	0.8	0.83	$7.2x10^{-01}$	0.67	$4.1x10^{-06*}$	- - -
COL4A3	rs11884770	2:228086920	T	25.1	28.0	0.85	$1.4x10^{-03}$	24.8	31.5	0.72	$1.3x10^{-03}$	30.8	33.7	0.87	$1.4x10^{-01}$	0.83	$6.6x10^{-06*}$	- - -
MUC1	rs4072037	1:155162067	C	45.3	47.6	0.89	$1.6x10^{-03}$	44.4	49.5	0.80	$5.8x10^{-04}$	42.3	43.8	0.94	$4.6x10^{-01}$	0.88	$9.6x10^{-06*}$	- - -

Common variants (minor allele frequency [MAF] $\geq$ 1%) in cases and in controls across all three populations with significant (P-value $< 1.24 \times 10^{-06}$) or suggestive ($P < 2.49 \times 10^{-05}$) associations with advanced AMD as discovered in the meta-analysis.

*Suggestive.

[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.

($P = 6.2 \times 10^{-12}$), and a strong signal for a common protective variant near *CTRB1* (rs8056814, $P = 1.8 \times 10^{-07}$).

The rare non-synonymous variant rs145732233 (A307V) in *PELI3* is associated with a decrease in risk of advanced AMD. *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 a 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 signalling transduction between the TLR3 and interferon regulatory factor 7 (IRF7) pathways, resulting in the downregulation of type I interferon expression (29). Individuals with the *PELI3* A307V mutation, therefore, may have less severe type I interferon response than individuals with normal pellino E3. Individuals with the *PELI3* A307V mutation also 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 4 (*TLR4*) (32) have been suspected to be related to AMD, these common associations were not replicated in studies of large cohorts (15–18). Through genotyping a large number of samples using the exome array, this study identified a novel, rare mutation in a gene that intermediates signals between TLRs and the innate immune pathway. Studies of the functional role of this protective mutation and possible mechanisms associated with AMD are warranted.

We identified a second protective rare variant significantly associated with a decrease in risk of advanced AMD. The missense SNP rs35274867 (*CFH* N1050Y) (OR = 0.76; $P = 6.2 \times 10^{-12}$) is located in *CFH*, a complement pathway gene that is strongly associated with AMD risk. An analysis of *CFH* N1050Y conditioned on other risk variants in *CFH* demonstrated that this SNP is independently associated with AMD (Table 1). Haplotype construction using the Boston cohort also shows that the protective allele for *CFH* N1050Y occurs independently of the risk allele for *CFH* R1210C (Table 4). Other studies have reported an association between AMD and both common and rare *CFH* variants (4–9,22,33,34). To our knowledge, this is the first report of a significant association between *CFH* N1050Y and AMD. This variant had a higher frequency in controls compared to cases in studies of systemic lupus erythematosus (35) and atypical hemolytic uremic syndrome (36), a disease that shares genetic influences with AMD. A descriptive report listed a 44 year-old patient with basal laminar drusen who carried this variant, but no analyses were conducted to evaluate or quantify an association (37).

CFH N1050Y is located in complement control protein module 18 (CCP18), a module that harbours multiple disease-associated

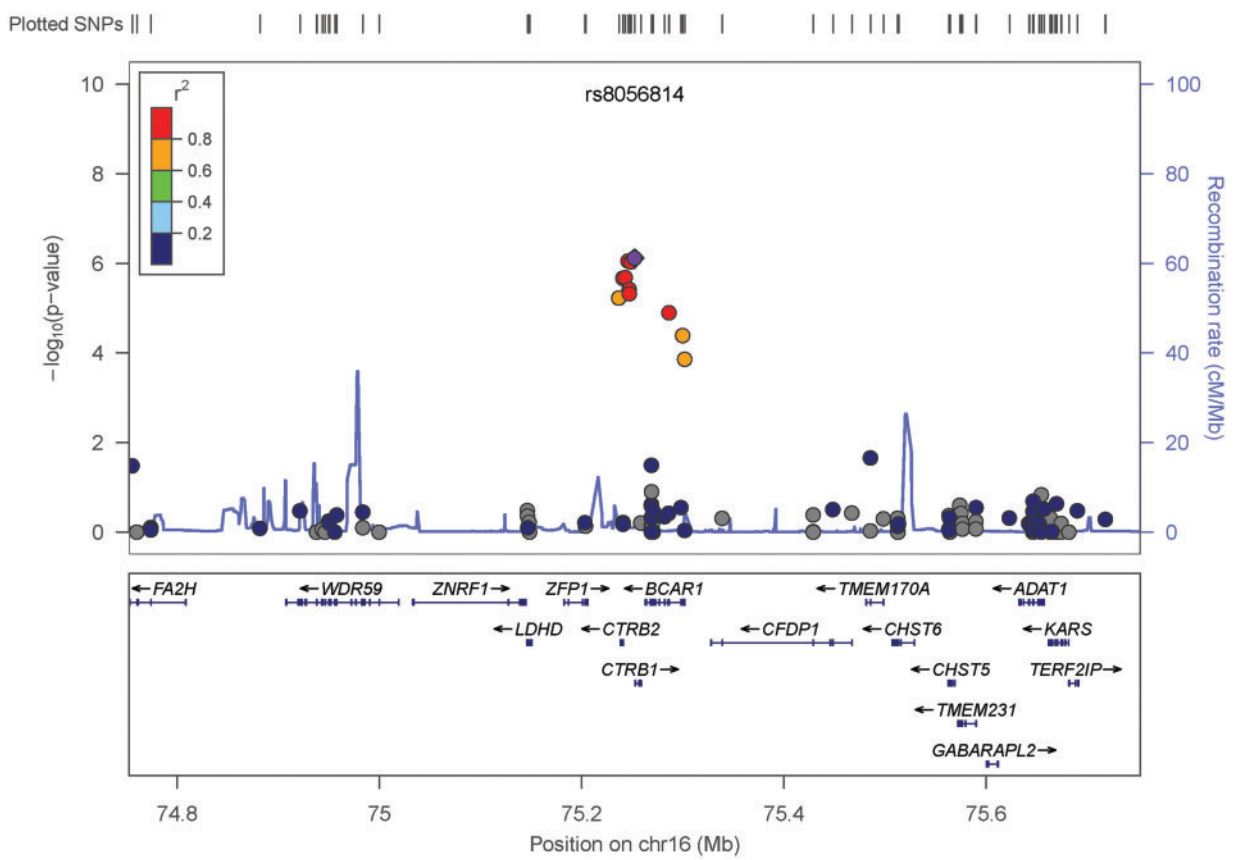

Figure 1. Zoomplots summarizing association results for the *CTRB1* locus. The regional association plot from analysis of common variants. Gene location is shown along the bottom of the graph, with observed $-\log_{10}$(*P*-value) along the left Y-axis and recombination rate along the right Y-axis. Each variant is plotted as a circle, colour-coded according to the extent of linkage disequilibrium. Variants with missing LD information are shown in grey. The index variant, rs8056814, is shown in purple.

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.

loci and may be part of an essential binding site (38). Alteration of binding domains of *CFH* could modify how the factor H protein regulates the alternative complement pathway. As this variant is associated with a lower 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.

Burden analyses detected a suggestive association signal in the *CFI* gene ($P_{Burden} = 3.2 \times 10^{-05}$, $P_{SKAT} = 1.0 \times 10^{-02}$), although not as strong as the association signal we recently identified in our sequencing study by simple burden tests (21). The disparity in signal strength may be attributed to the fact that there were only 12 *CFI* rare variants detected through this exome array, compared to the 59 *CFI* rare variants we previously detected by targeted sequencing. Our results exemplify the value of employing a combination of genotyping and sequencing platforms for gene discovery.

Our analyses of common variants identified an association between AMD and the protective variant rs8056814 in *CTRB1*. This SNP is approximately 330kb away from rs8053796 in contactin associated protein-like 4 (*CNTNAP4*), where we detected a suggestive association signal ($P = 1.7 \times 10^{-05}$) in our previous meta-GWAS study (17). Other studies have reported nominal associations between rs8056814 and AMD in the same direction of effect, and an independent, concurrent study reported a significant association between rs8056814 and AMD (18,39–41). Compared with some previous studies, rs8056814 was directly genotyped in our analysis rather than imputed, which enabled us to assess the association at this locus more accurately and with more power. The effect of this locus may have been subsequently underestimated in studies using imputed datasets, as the linkage disequilibrium structure may not have perfectly captured its genotype information.

The rs8056814 variant is located 557bp upstream of *CTRB1* and encodes chymotrypsinogen, a serine protease that is secreted into the gastrointestinal tract and activated by proteolytic cleavage with trypsin. Querying the GTEx database, we found that rs8056814 is a cis-expression quantitative trait locus (eQTL) for the downstream gene *BCAR1* in whole blood ($P = 1.5 \times 10^{-07}$, $\beta = 0.37$) (data were obtained from the GTEx Portal and dbGaP accession number phs000424.v6.p1) (42). CHIPseq results from the ENCODE project suggest that rs8056814 lies in the promoter region of *CTRB1* in a hepatic carcinoma cell line and in an enhancer region in blood cell lines, potentially explaining the eQTL association with *BCAR1* in blood (43). Eye tissue was not assessed as part of the GTEx and ENCODE projects. The GeneAtlas U133A BioGPS database reported some mRNA expression of *CTRB1* in the retina, but the level is similar to many other tissues in the model. Retinal expression of *CTRB1* was only evaluated in one of three expression methods and requires further investigation to determine the role of the rs8056814 variant in ocular tissues (44).

Degradation of the ECM pathway has been determined to contribute to the pathogenesis of AMD, as evidenced by an association with *COL8A1* and potentially other genes, including *COL4A3* suggested herein (17,18). *CTRB1* is related to this degradation and is in a region of chromosome 16 also reported to be associated with variation in HDL levels (17,45,46). 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) is associated with higher HDL levels and also a protective effect on the risk of AMD (15,16). Considering the effects of other variants in this pathway (46), it is possible that the minor allele of rs8056814 could influence metabolic changes in HDL levels through modulation of *CTRB1* expression levels.

While the Illumina Infinium HumanExome BeadChip may serve as an alternative to whole exome sequencing for an unbiased assessment of functional variants, it does not provide complete coverage of all functional variants in all genes. 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 (21,24). The lack of information on those rare variants could weaken the association signals in the gene-based test, especially in the scenario where many rare variants in a gene are 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 additional investigation, including common variants in *COL4A3*, *APOH* and *VTN*, 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* we discovered by targeted sequencing. The new genetic loci associated with AMD suggest that genes in other branches of the innate immune pathway and additional genes in the ECM and HDL pathways may also be involved in the aetiology of AMD. As knowledge of the genetic architecture of AMD expands, new variants may enhance predictive models (50,51), and could lead to the development of new therapeutic targets and more directed and personalized management of this disease.

Materials and Methods

Case-control definitions

All individuals were evaluated by a board-certified ophthalmologist who conducted ocular examinations including visual acuity measurements, dilated slit-lamp biomicroscopy and stereoscopic colour 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). Cases were defined as patients who 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 were individuals who 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 cases 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). Individuals from all cohorts were self-reported white individuals of European descent (Supplementary Material, Table S5).

We included genotype data for 2,909 shared controls genotyped at the Broad Institute (29,58–61). 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-examined with five outlier removal iterations in EIGENSTRAT to ensure that samples were appropriately matched according to their ethnicity (Supplementary Material, Fig. S8).

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 our AMDGENE consortium meta-GWAS study (18). Genotyping of the Boston, French and Finnish cohort samples was performed at the Johns Hopkins Genotyping Core Laboratory. Shared control samples were genotyped at the Broad Institute using the same genotyping platform and custom content. 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. 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 allele frequencies significantly different between the examined controls genotyped at the Johns Hopkins Core Laboratory and the shared controls genotyped at the Broad Institute ($P < 10^{-03}$). We identified 16,008 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 proportion identity-by-descent estimates (PIHAT values) using the remaining 15,671 ancestry informative markers (64). We identified related pairs of sequenced individuals with PIHAT > 0.2, and removed one individual of each pair 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 in each cohort, we tested for associations assuming an additive genetic model using logistic regression adjusting for the first 10 principle components of the EIGENSTRAT analysis. The summary data of all cohorts were then meta-analyzed using a random effects model with METAL (65). In addition to testing each variant for an association with overall advanced AMD, we conducted separate subtype analyses and tested each variant for an association with each of 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 21 known AMD loci (Supplementary Material, Table S3) by adjusting for the genotype of the adjacent known variant. Significance ($P < 5 \times 10^{-08}$) and suggestive ($P < 1 \times 10^{-05}$) thresholds were used to evaluate the association signals of common variants.

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 modified Fisher's exact test to calculate a one-tailed exact *P*-value across multiple case-control cohorts based on the exact hypergeometric probability of observing each particular arrangement of the case-control data stratified by each cohort, assuming the given marginal totals in each cohort. We performed analyses on 57,101 variants with a MAF <1%. We applied the same statistical framework on data further stratified by the genotypes of nearby known common variants, in addition to country of sample collection, to calculate the *P*-values for the conditional analysis of 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. Significance ($P < 5 \times 10^{-08}$) and suggestive ($P < 1 \times 10^{-05}$) thresholds were used to evaluate the association signals of rare variants. To further ensure that the significant rare associations were not due to any hidden confounding factors or driven by case-control imbalance among the cohorts, we performed Firth logistic regression analysis adjusting for the top 10 principal components on the top significant rare associations after excluding the additional Finnish controls (>20,000 from the National FINRISK Study) using the logistf package in R. (Summary data for common and rare loci are available in the Supplementary Material, Tables S6–S7.)

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 non-synonymous, nonsense, or splice-site variants with MAF between 1 and 0.03% in case or control groups. We excluded extremely rare variants (minor allele count < 5, or MAF < 0.03%) from this test. These variants were located in 12,122 genes. Each gene contained 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 significance, we applied a Bonferroni corrected significance threshold of $P < 2.06 \times 10^{-6}$ ($P = 0.05/12{,}122 \times 2$ gene-based tests).

Haplotype construction and frequency calculation of the *CFH* region were performed using the samples in the Boston cohort. Seven SNPs in *CFH* were used in the haplotype construction including the rare *CFH* SNPs R1210C and N1050Y. 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).

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 $P < 5 \times 10^{-08}$.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We appreciate the assistance of Rachel E. Silver, MPH and Mercedes B. Villalonga, BA (Ophthalmic Epidemiology and Genetics Service, Tufts Medical Center) with the critical review and revision of this manuscript, as well as their continued oversight of the Boston study cohort. We thank the numerous ophthalmologists throughout the United States who 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.

Conflict of Interest statement. JMS: research grant - Novartis.

Funding

Components of this project were supported by the National Institutes of Health [R01-EY011309, 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 genotyping was funded by 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.

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. USA*, **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.
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. USA*, **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. USA*, **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. *Nature*, **17,** 536:285–291.
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 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 Res. Ther.*, **13**, R206.
36. 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.
37. 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.
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. 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.
43. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
44. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA*, **101**, 6062–6067.
45. GeneCards. Available at: http://www.genecards.org/cgi-bin/carddisp.pl?gene=CTRB1. (Accessed: 27th July 2016).
46. 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.
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. *Eur. J. 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. *Int. J. Epidemiol.*, **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.
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.