Rare variants in *CFI, C3* and *C9* are associated with high risk of advanced age-related macular degeneration

Johanna M. Seddon[*,1,2,3], **Yi Yu**[1], **Elizabeth C. Miller**[4], **Robyn Reynolds**[1], **Perciliz L. Tan**[5,6,7], **Sivakumar Gowrisankar**[8], **Jacqueline I. Goldstein**[9,10], **Michael Triebwasser**[4], **Holly E. Anderson**[11], **Jennyfer Zerbib**[12], **David Kavanagh**[11], **Eric Souied**[12], **Nicholas Katsanis**[5,6,7], **Mark J. Daly**[9,10], **John Atkinson**[4], and **Soumya Raychaudhuri**[*,8,9,13,14,15]

[1]Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, Massachusetts, USA.

[2]Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts, USA.

[3]Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, Massachusetts, USA.

[4]Division of Rheumatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

[5]Center for Human Disease Modeling, Duke University, Durham, North Carolina, USA.

[6]Departments of Cell Biology, Duke University, Durham, North Carolina, USA.

[7]Pediatrics, Duke University, Durham, North Carolina, USA.

[8]Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA.

[9]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA

[10]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

[11]Institute of Genetic Medicine, Newcastle University, International Centre for Life, Newcastle upon Tyne, UK

[12]Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France

[13]Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, USA.

*Please send correspondence to: 800 Washington St. #450 Tufts Medical Center Boston, Massachusetts 02111 United States of America jseddon@tuftsmedicalcenter.org 617-636-9000 (tel) 617-636-5844 (fax). *77 Avenue Louis Pasteur Harvard New Research Building, Suite 250D Boston, Massachusetts 02446 United States of America soumya@broadinstitute.org 617-525-4484 (tel) 617-525-4488 (fax) .

[14]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA.

[15]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

Abstract

To define the role of rare variants in advanced age-related macular degeneration (AMD) risk, we sequenced the exons of 681 genes within AMD-associated loci and pathways in 2,493 cases and controls. We first tested each gene for increased or decreased burden of rare variants in cases compared to controls. We found that 7.8% of AMD cases compared to 2.3% of controls are carriers of rare missense *CFI* variants (OR=3.6, p=2×10^{-8}). There was a predominance of dysfunctional variants in cases compared to controls. We then tested individual variants for association to disease. We observed significant association with rare missense alleles outside *CFI*. Genotyping in 5,115 independent samples confirmed associations to AMD with a K155Q allele in *C3* (replication p=3.5×10^{-5}, OR=2.8; joint p=5.2×10^{-9}, *OR*=3.8) and a P167S allele in *C9* (replication p=2.4×10^{-5}, OR=2.2; joint p=6.5×10^{-7}, *OR*=2.2). Finally, we show that the 155Q allele in C3 results in resistance to proteolytic inactivation by CFH and CFI. These results implicate loss of C3 protein regulation and excessive alternative complement activation in AMD pathogenesis, thus informing both the direction of effect and mechanistic underpinnings of this disorder.

Age-related macular degeneration (AMD) leads to irreversible vision loss in individuals over 55 years of age[1-3]. Most known genetic AMD variants are common, without established disease mechanisms[4]. To date, highly penetrant missense variants for AMD are only known in *CFH*[5]. We sought to identify additional rare missense risk variants, with the goal of identifying specific disease mechanisms.

We applied targeted next-generation sequencing to coding regions of candidate genes within AMD loci and pathways (**Online Methods**, Supplementary Figure 1). To maximize power, we prioritized cases with severe disease, familial disease or early age of onset; as well as older controls without drusen. We sequenced 681 genes to >20X coverage for a median of 95.6% of the targeted region (Supplementary Figure 2). After calling variants with GATK[6] and applying strict quality control, we obtained genotype data for 1,676 cases, 745 controls, and 36 siblings with discordant disease status (*n*=2,493). Our data had <0.92% missing genotypes for each allele and <0.54% missing genotypes for each individual. We observed effect sizes comparable to published values for known AMD risk variants or proxies (Supplementary Table 1). Many variants discovered in our sequencing study were not genotyped or imputed in recent AMD GWAS (Supplementary Figure 3).

We compared sequence-based genotypes to genotypes from the HumanExome BeadChip™ array (i.e. exome-chip)[7,8]. We observed 99.97% concordance for 2,426 missense SNPs that were captured by both sequencing and exome-chip with allele frequencies ≥ 0.001 (**Online Methods**). Allelic dosages were almost perfectly correlated (*r*>0.99) for 96.5% of common variants (*f*>0.01) and 93.0% of rare variants (Supplementary Figure 4); only 0.2% and 3.6% of common and rare SNPs, respectively had modestly correlated dosages (*r*<0.9).

With the sequencing data, we tested if any of 681 genes had a higher burden of rare variants in cases or in controls. We selected only rare variants (f<0.01) within targeted genes that altered coding sequences (missense, nonsense, read-through, or splice variants, n=15,762 SNPs). We used a simple burden test to assess whether the proportion of case individuals who carried at least one rare variant was in excess of chance (**Online Methods**, Supplementary Table 2A)[9]. We similarly tested increased burden in controls. To interpret statistical significance, we applied a Bonferroni-corrected significance threshold of p<3.6×10^{-5} (=0.05/(681 genes × 2 tests)). We saw a significantly increased burden in cases in only one gene, *CFI* (exact one-tailed p=1.6×10^{-8}, *OR*=3.57, Figure 1A); 7.8% of cases had a rare coding variant compared to 2.3% of controls. In contrast, using the C-alpha test[10] we found no evidence of any genes with alleles acting in opposite directions (Supplementary Table 2A).

The enrichment of rare coding variants in *CFI* is likely not due to confounding by population stratification, sequencing artifact, age or gender. First, if the enrichment was due to stratification, then other genes would also demonstrate enrichment for rare variants; however, burden tests for other genes did not exceed expectation due to chance (Figure 1A). Second, stratification would also drive enrichment of synonymous variants; however, we observed no significant case enrichment for rare synonymous variants in *CFI* (p=0.89), or in any of the 681 genes (p>0.001, Supplementary Table 2A, Supplementary Table 3). Third, to adjust for case-control stratification, we calculated principal components based on ancestry-informative markers in a subset of 1,558 cases and 683 controls with exome-chip genome-wide genotyping data. We found that *CFI* rare variant carriers did not cluster along the top two principal components (Supplementary Figure 5). Furthermore, adjusting for the top five principal components in a logistic regression framework effected no change in the case enrichment of rare *CFI* coding variants (p=5.0×10^{-8}, *OR*=3.6, versus p=1.1×10^{-7}, *OR*=3.5 after adjustment). To rule out sequencing artifact, we confirmed 137/140 assayed rare variant events with Sanger sequencing (see Supplementary Figure 6 and **Online Methods**). The observed enrichment remained significant after adjusting for age and gender (p=6.7×10^{-9}, *OR*=3.7 versus p=1.1×10^{-8}, *OR*=3.6 after adjustment). We note that since extreme phenotypic samples were prioritized for sequencing, independent replication in prospectively ascertained samples will ultimately be necessary to accurately estimate relative enrichment of rare *CFI* variants within the general population.

The case-enrichment of rare coding variants in *CFI* was independent of the nearby common risk allele, rs4698775[4,11]. When we stratified our samples for rs4698775 genotypes, we observed consistent *CFI* rare coding variant enrichment across all three rs4698775 genotypes (p=1.7×10^{-8}, Figure 1B). Moreover, the risk conferred by the common rs4698775 allele was independent of rare variant carrier status (p<0.04, allelic *OR*=1.15), arguing against a "synthetic association".

We then used PolyPhen-2[12] to classify the 59 separate *CFI* nucleotide variants into four categories based on their potential predicted impact (1) loss of function (nonsense or splice variants)[13], (2) probably damaging, (3) possibly damaging and (4) benign (Figure 2A, Supplementary Table 4). Loss of protein function variants (three nonsense and one splice variant) were the most skewed and exclusively found in cases (7 to 0, Figure 2B). Probably

damaging variants were more skewed towards cases (41 cases to 3 controls) than possibly damaging variants (16 to 3) and benign variants (70 to 12). Indeed, probably damaging or loss of function variants conferred high risk of disease (OR=7.5, p=1.3×10^{-5}). Further, some of the variants classified as benign are potentially pathogenic, and inaccurately classified by PolyPhen-2[14].

CFI encodes complement factor I, that consists of a heavy chain and a light chain, with a catalytic serine protease domain which cleaves the C3 protein. There is a greater burden of rare variants in the catalytic light chain (OR=4.85, p=2.2×10^{-6}) than in the heavy chain (OR=2.63, p=0.0012). We observed no evidence of association with AMD for any of the individual *CFI* variants (p>0.02), possibly due to their low frequency.

Remarkably, many of these *CFI* rare variants have been found in patients with atypical hemolytic uremic syndrome (aHUS), including P50A[15], G119R[15], A240G[16], G261D[17], R317W[16], I340T[18], Y369S[16,19], D403N[15], I416L[19], Y459S[15], R474X[20], and P553S[15]. These variants likely cause aHUS by causing haploinsufficiency and thus mitigating the ability of CFI to cleave and thereby deactivate C3b. The R317W and I340T variants result in CFI functional deficiency[21], while P50A, I416L, and R474X produce a CFI protein quantitative deficiency[15]. Despite extensive study, no functional deficiency for G261D has been identified[17].

We also tested rare variants individually for association with AMD (Supplementary Figure 1). This study was not powered to detect association for variants observed <5 times; thus we excluded those extremely rare SNPs from this analysis. We tested 1,824 variants within the 681 sequenced genes that had an allele frequency of <1% in cases or controls (**Online Methods**, Supplementary Table 2B). To interpret statistical significance, we applied a Bonferroni corrected significance threshold of p<1.4×10^{-5} (=0.05/(1,824 variants × 2 tests)).

We identified five risk and 15 protective variants that demonstrated nominal evidence of association in discovery (exact 1-tailed p<0.01, Supplementary Table 5). Four of these variants were within or near *CFH*, including the previously reported *CFH* R1210C risk variant (p=0.0012)[5]. In addition, we observed association to the *CFH* N1050Y, *CFHR2* Y264C and *CFHR5* G278S protective variants. After phasing common variants within the *CFH* locus, we observed that the rare associated variants were in LD with *CFH* haplotypes (Supplementary Figure 7). We were uncertain whether these associations represented unique effects or were tagging *CFH* haplotypes with large effects, driven by the Y402H and rs10737680 common risk alleles[22-24].

We evaluated 11 of the 16 variants outside of the *CFH* locus in an independent set of 2,227 cases and 2,888 controls from Boston, Baltimore, and France genotyped with exome-chip or Taqman (Supplementary Table 6). To increase power, we incorporated shared controls in this replication study (**Online Methods**); these controls may result in decreased allelic effect sizes since some may ultimately develop advanced AMD. We observed independent evidence of association for two variants in replication (p<0.0031=0.05/16) and joint analysis (p<1.4×10^{-5}) : rs147859257 (K155Q) variant in *C3* (exact 1-tailed p=4.8×10^{-6} in discovery) and the rs34882957 (P167S) variant in *C9* (p= 2.3×10^{-3} in discovery, Table 1,

Table 2, Supplementary Table 5). Both variants had perfectly concordant sequence-based and exome-chip genotypes (r=1). To assess if either association could be related to population stratification, we calculated principal components for a subset of samples (n=2,241), as described above. Individuals that were carriers of either variant did not cluster along the first two principal components (Supplementary Figures 8 and 9). Also, adjusting for the top five principal components effected no change in the K155Q effect (p=2.7×10^{-5}, OR=16.1 versus p=2.0×10^{-6}, OR=16.7, after adjustment) or in the P167S effect (p=0.053, OR=2.0 versus p=0.053, OR=2.0, after adjustment). These associations retained significance after adjusting for age and gender (p=1.5×10^{-5}, OR=9.5 versus p=1.5×10^{-5}, OR=9.0 for K155Q after adjustment; p=0.007, OR=2.3 versus p=0.01, OR=2.2 for P167S after adjustment).

The *C3* K155Q (position 133 excluding the signal peptide) variant demonstrated compelling evidence of association in replication (p=3.5×10^{-5}, OR=2.8) and was highly significant with a large effect size in joint analysis with discovery samples (p=5.2×10^{-9}, OR=3.8; Table 1). While the estimated effects for K155Q in discovery and replication have overlapping confidence intervals, the effect in discovery is larger, perhaps related to the "winners curse"[25] or inclusion of shared controls in replication. This risk is independent from a previously-reported common risk variant, R102G (rs2230199)[26,27]. In a subset of samples with genotypes for R102G and K155Q (n= 6,935, excluding Baltimore samples), we observed an increased significance of the K155Q effect when we stratified individuals on R102G genotypes (from p=5.4×10^{-9} to p=6.2×10^{-10}, Table 1). The K155Q risk variant is in phase with the protective 102R (rs2230199-G/102-R) allele. The *C9* P167S variant also demonstrated convincing association in replication (p=2.4×10^{-5}, OR=2.2;) and in joint analysis (p=6.5×10^{-7}, OR=2.2; Table 2). While the P167S variant is predicted with high confidence to be a probably damaging variant by PolyPhen-2[12], there are no reported functional data. We note one other independent nominal association has been previously reported in *C9*[28]. This expands the repertoire of AMD genes in the complement cascade, specifically implicating the membrane attack complex (C5-C9) formed downstream of C3 activation in the alternative complement pathway activation.

We note that K155Q is on the surface of C3, close to the binding site for CFH, (Supplementary Figure 10)[29] which is a cofactor for CFI-mediated cleavage of C3. We therefore tested the hypothesis that the K155Q substitution in C3 might confer resistance to its inactivation by CFI using a previously-described experimental strategy for evaluating aHUS mutations[30]. Importantly, we observed significantly reduced cleavage of 155Q compared to 155K C3 in fluid phase cofactor assays (Figure 3A and B). We also observed significantly reduced binding to CFH by 155Q in surface plasmon resonance (~60%, Supplementary Figure 11A) and a trend toward reduced binding with ELISA (Supplementary Figure 11B). In contrast, MCP cofactor activity was not reduced by K155Q (Supplementary Figure 12A-C).

We conclude that *C3* K155Q primarily impairs C3b inactivation by CFI with bound CFH, resulting in increased C3 convertase formation and feedback amplification of the alternative pathway (Supplementary Figure 13). The effect of K155Q mirrors that of the highly penetrant R1210C substitution in CFH, which disrupts CFH binding to C3b[31-33]. The

burden of *CFI* mutations, particularly the loss of function mutations in the catalytic domain of the protein, along with the large effect of the K155Q and R1210C variants underscore the critical role for impaired alternative pathway regulation by dysfunction of the C3b-CFH-CFI tri-molecular interaction. These observations parallel mounting evidence that common AMD risk alleles might also lead to increased alternative pathway activity[34-36]. These results suggest the potential of agents that prevent C3 activation, or enhance CFI or CFH, to reduce the risk or progression of disease.

Strikingly, our study did not identify rare variants outside the complement pathway, despite the large set of genes queried within other AMD loci and genes involved in angiogenesis, lipid metabolism and extracellular matrix[4,10,37,38]. One possibility to explore as sequencing studies scale is that rare coding variants with even more modest effect sizes than those reported here modulate AMD risk within those other genes as well.

ONLINE METHODS

Study Sample Descriptions

Case Definitions—Board certified ophthalmologists evaluated all case subjects and matched (non-shared) control subjects. We either (1) clinically evaluated with visual acuity measurements, dilated slit lamp biomicroscopy and stereoscopic color fundus photography or (2) reviewed full ophthalmologic medical records and images. Case patients had either geographic atrophy (advanced dry AMD) or neovascular disease (wet AMD) according to the Clinical Age-Related Maculopathy Grading System (CARMS) stages 4 and 5[39]. Controls were also examined and had no signs of intermediate or advanced macular degeneration in either eye and absence of bilateral early AMD. All Boston and France controls and most Baltimore controls (>80%) were 60 years old.

Boston—Subjects were recruited through ongoing AMD study protocols[2,11,24,40-42]. After quality control, these samples ultimately included 2,422 unrelated cases and 1,287 unrelated controls, and also 49 discordant sib-pairs. We applied targeted sequencing to a subset of these samples, including 1,676 unrelated cases and 745 unrelated controls, as well as 36 siblings with discordant case status (see below) passing quality control. We applied the Illumina Infinium HumanExome genotyping array (see below) to an independent set of 746 cases, 542 controls, and 13 siblings with discordant case status for replication passing stringent quality control. To assess concordance, we successfully genotyped a subset of 1,558 sequenced cases and 683 sequenced controls with the Illumina HumanExome genotyping array. Sequenced controls had a mean age of 77.2, while sequenced cases had a mean age of 75.6 (see Supplementary Figure 14).

France—We recruited AMD cases and controls at the Hopital Intercommunal de Creteil (FR-CRET), Creteil, France, as previously described[10,37], who were self-described white individuals of European descent. We genotyped all samples with the Illumina HumanExome genotyping array. These samples included 953 unrelated cases and 203 unrelated controls after quality control.

Baltimore—Unrelated subjects were recruited at Johns Hopkins Hospital in Baltimore, MD as previously described[10,38,43,44], and were self-described white individuals of European descent. After quality control, we obtained genotype data from 516 cases and 163 controls for selected SNPs with TaqMan (see below).

Shared controls—We utilized a collection of shared control samples from four separate studies, broadly consented for medical use, genotyped at the Broad Institute. In aggregate, these samples included 2,466 samples passing quality control from individuals who were not evaluated for retinal diseases, Caucasian, and unrelated. The samples included control individuals recruited for the 1000 genomes project (n=448)[45], the International Severe Adverse Event Consortium (iSAEC, n=709)[46], the National Institutes of Mental Health controls (n=1,054)[47], and the Prospective Registry in IBD Study at MGH (PRISM, n=255)[48]. Ultimately, in the final analysis we selectively added 1,491 of these shared controls to the Boston collection and 476 of these shared controls to the France collection (see below).

Targeted Sequencing

Gene Selection—We first selected all genes within 19 genomic regions defined by common SNPs which have been associated with AMD and obtained genome-wide significance in a recent large meta-analysis[4]. Then, we selected genes closest to common SNPs with nominal significance ($p<10^{-4}$) in a smaller previous meta-analysis[37]. Next, we selected genes involved in pathways implicated in AMD or other retinal diseases, including complement genes, angiogenesis genes, genes involved in the structure of retinal pigment epithelium (RPE) and photoreceptors, HDL metabolism genes, genes involved in inflammation and oxidation pathways, and genes in the extracellular and collagen matrix pathways. Finally, we included genes reported to be associated with related diseases, including Stargardt disease, Best disease or vitelliform macular dystrophy, Alzheimer's disease, and atypical hemolytic uremic syndrome (aHUS).

Targets capture and re-sequencing—We sequenced samples at PerkinElmer, Inc. We designed a custom Agilent SureSelectXT Kit to capture genomic sequences of coding exons, splice junctions, 5′ UTR and 3′ UTR regions in 765 selected genes with indexing barcodes for each sample. We sequenced a total target length of 5.28Mb including 1.76 MB of coding exons. We isolated the hybridized library fragments, quantitated by qPCR and sequenced, and then sequenced paired-end reads with the Illumina HiSeq2000™ sequencing platform. We required sequencing data for each sample to have over 10X coverage at greater than 90% targeted regions and over 20X coverage at greater than 80% targeted regions.

Read mapping, variant detection and annotation—We aligned sequence reads in each individual to the human reference genome (NCBI build 37.3, hg19) with BWA (v0.59)[49]. We called the consensus genotypes in the target regions with The Genome Analysis Toolkit (GATK, v2.18) with the workflow and parameters recommended in the best practice variant detection with GATK v4[6,50]. We applied GATK duplicate removal, indel realignment, base quality score recalibration, and performed multi-sample SNP and indel discovery and genotyping across all samples simultaneously using variant quality score

recalibration (VQSR). Other than high quality variants assigned "PASS" by VQSR, we also included only those variants in lower tranches with truth sensitivity between 99.0-100 that were also separately recorded in the exome sequencing project (ESP) database of 6500 samples[51]. We annotated variants with snpEff (v2.05)[52].

Quality Control—We further excluded SNPs failing Hardy-Weinberg test in controls ($p<10^{-6}$). We also excluded alleles from further analysis that had high missing genotype data (>1%), likely due to systematic low coverage or difficulty mapping reads across many samples. We also excluded samples with high missing genotype data (>1%) for common alleles with >1% frequency in our data set. For burden testing, we only tested those genes on autosomes that obtained >10X coverage at an average of >90% of the targeted region and had rare coding variants (*n*=681).

Validation Sanger Sequencing

We sequenced the *CFI* gene in individuals that were carriers of rare *CFI* variants at Beckman Coulter Genomics. We designed primers (see Supplementary Table 7) to produce 400-600 bp amplicons that were at least 50 bp from the intron/exon boundary was included to assure high quality sequence for splice sites and the entire exon sequence. We sequenced forward and reverse PCR primers via capillary electrophoresis using ABI Prism 3730xl DNA analyzers (Applied Biosystems, Foster City CA). We assembled chromatograms using CONSED and detected polymorphisms with polyphred 5.04 (University of Washington).

Array-based genotyping of coding variants

We genotyped the France and Boston sample sets with the Illumina Infinium HumanExome BeadChip (v1.0), which provides coverage of over 240,000 functional exonic variants selected from >12,000 whole exome and whole genome sequences. In addition, we customized our assay by adding 3,214 SNPs in candidate AMD genes. We conducted genotyping at the John Hopkins Genotyping Core Laboratory. We genotyped shared control samples separately at the Broad Institute using the Illumina HumanExome v1.0, v1.1 and v1.1+custom content SNP arrays.

We called genotypes using Illumina's GenomeStudio software and then zCall[7]. We required that samples have <2% missing genotype calls for common variants (MAF>5%) before applying zCall. Then after applying zCall we removed duplicate SNPs, monomorphic SNPs, SNPs with a low call rate (<98%), and SNPs failing Hardy-Weinberg ($p<10^{-6}$). We merged genotype calls from the four shared control cohorts and the AMD cohort by only including SNPs that passed quality control in all 5 cohorts and passed Hardy-Weinberg test ($p>10^{-6}$) across all samples.

Taqman genotyping of selected SNPs

We genotyped Baltimore samples at the Duke University Center for Human Disease Modeling using a custom made TaqMan genotyping assay by Applied Biosystems and with the ABI 7900 Real-Time PCR system (see Supplementary Table 7).

Statistical Analyses

Removing related samples—For sequenced samples, we assessed relatedness by calculating genome-wide proportion identity-by-descent estimates (PIHAT values) using common (f>5%) SNPs pruned to insure independence with the --indep option in PLINK with default parameters (VIF=2, window size = 50 SNPs)[53]. We identified pairs of sequenced individuals with PIHAT > 0.2, who were not known discordant sibling pairs, and removed one of those individuals from the analysis. We similarly estimated PIHAT for all replication samples genotyped by exome-chip. We removed any individual in replication with PIHAT > 0.2 to another individual or to sequenced discovery samples. We also identified pairs of individuals in replication samples with PIHAT > 0.2, that were not known discordant sibling pair, and removed one of those individuals from the analysis.

Assessing concordance of array-based and sequencing based genotypes— Because rare SNPs are challenging to genotype with array-based platforms, we selected only those alleles with minor allele counts of 5 (i.e. allele frequency >0.1%). We also selected only those SNPs with 0% missingness in the array data to ensure the highest possible accuracy for array-based genotypes. To assess concordance we calculated a simple concordance between genotype dosages across individuals with the two different platforms. This correlation-based metric is comparable across allele frequency spectrums.

Ancestry informative markers—For samples genotyped on the exome chip we identified 16,008 ancestry informative markers. To identify these SNPs, we selected SNPs on the exome-chip with allele frequencies (f>5%), and excluded regions in the *CFH* locus (chr1, 195.5-197.5 MB in HG19 coordinates), the *ARMS2* locus (chr10, 123-125 MB), and the Major Histocompatability Complex locus (chr 6, 25-35 MB) loci. We then pruned the resulting set of SNPs using the --indep option in PLINK with default parameters (VIF=2, window size = 50 SNPs)[53].

Combining shared controls—For samples genotyped with exome-chip in the France and Boston data sets, we included shared controls to increase power. In order to mitigate the potential effects of population stratification in these replication data sets we included shared controls into each collection by matching on case ancestry. First we applied EIGENSTRAT to these SNP genotypes to calculate principal components to match samples in replication samples together with shared controls[54]. Then we used the first 5 principal components to calculate Euclidean distances between samples in the Boston and France cohorts and shared control samples. Finally, we randomly selected individual case samples in these cohorts and assigned the nearest unassigned shared controls to the selected case's cohort. Shared controls could only be assigned to one of two cohorts. We added about one shared control for every two cases to the France collection, so that in total we added 476 shared controls. We also added about two shared controls for every one case to the Boston replication collection, so that in total we added 1,491 shared controls. The resulting expanded Boston and France cohorts had minimal evidence of population stratification (λ_{gc} = 1.04 and λ_{gc} = 1.06 respectively for ancestry informative markers).

Statistical framework for association testing—Since asymptotic statistics can be inaccurate for rare variants, we utilized exact statistics instead to test association for individual variants and also for sets of variants in genes (burden tests). We used a strategy similar to Raychaudhuri et al[5].

For single strata case-control sample collections we used a 2×2 Fisher's exact test to calculate a one-tailed *p*-value. Assuming a single strata that there are a total of *N* individuals, of whom n_{case} are cases and $n_{variant}$ are carriers for the variant, we calculate the one-tailed significance of observing $n_{variant,case}$ individuals who have the variant and have advanced AMD as follows:

$$p_{\text{case-control}}\left(n_{\text{variant,case}}\right) = \sum_{n_{\text{variant,case}} \leq x \leq n_{\text{case}}} hg\left(x, N, n_{case}, n_{\text{variant}}\right)$$

where *hg* is the hypergeometric probability distribution assuming that there are $n_{variant}$ draws from a total of *N* samples, of which *x* of a total of n_{case} are drawn.

If multiple case-control strata are present, for example if we are stratifying genotypes of a common variant or combining multiple case-control collections together, we expand the above strategy to calculate an exact *p*-value. Assume that we observe a total of $n_{variant,case}$ carriers who are affected across all strata then, we can calculate significance as follows:

$$\begin{aligned} &p_{\text{stratified,case-control}}\left(n_{\text{variant,case}}\right) \\ &= \sum_{\text{sum}(x_1 \ldots x_j) \geq n_{\text{variant,case}}} hg\left(x_1, N_1, n_{1,\text{case}}, n_{1,\text{variant}}\right) \cdot \ldots \cdot hg\left(x_j, N_j, n_{j,\text{case}}, n_{j,\text{variant}}\right) \end{aligned}$$

Here, for each strata *j* we have separate numbers of total individuals (N_j), separate numbers of individuals who are cases ($n_{j,case}$), and heterozygote individuals ($n_{j,variant}$). So, we calculate the hypergeometric probability for each individual strata for all the possible counts that would result in an equal or greater than $n_{variant,case}$ total number of heterozygotes associated with advanced AMD, and total these probabilities together to obtain a *p*-value.

For discordant siblings we calculated statistical significance using the binomial distribution. For a given variant, we consider only those pairs of siblings with discordant genotype for the rare variant. The probability under the null that the affected sibling will have the variant is 0.5. We assign each discordant sibling pair a score, s_i, which is 1 if the affected sibling has the rare variant or 0 if the unaffected sibling has the rare variant. We obtain an aggregate score by summing over all independent siblings:

$$s_{\text{siblings}} = \sum_i s_i$$

Under the null hypothesis, the aggregate score should be distributed according to a binomial distribution. So if there are a total of $n_{siblings}$ we can calculate $p_{sibling}$, the one-tailed significance:

$$p_{\text{siblings}}\left(s_{\text{siblings}}\right) = \sum_{x > s_{\text{siblings}}} \text{binomial}\left(x, n_{\text{siblings}}, 0.5\right)$$

where the function *binomial* represents the binomial distribution for *x* successful draws out of $n_{siblings}$ each with a 0.5 probability.

In order to calculate an aggregate meta-analysis we define a score, $s_{aggregate}$ which is the total of $s_{variant,case}$ across all strata and siblings. We can calculate the probability of obtaining the score *s* or a more significant score to determine the exact one-tailed *p*-value:

$$p_{\text{meta}}\left(s_{\text{aggregate}}\right) = \sum_{s_{\text{aggregate}} \le \text{sum}(x_1 \ldots x_j) + y} \text{binomial}\left(y, n_{\text{siblings}}, 0.5\right) \cdot \prod_j hg\left(x_j, N_j, n_{j,\text{case}}, n_{j,\text{variant}}\right)$$

Association Testing—We tested nonsynonymous, splice acceptor-site and donor-site, start lost, stop gained and stop lost variants for association, that were rare variants (*f* 0.01 in either cases or controls). We only included rare variants with a minor allele count   5 in our dataset. To conduct stratified testing for other nearby common SNPs, we further subdivided strata based on common variant genotypes.

Burden testing—To test for burden, we identified those variants that altered coding sequences (missense, altered splice acceptor-sites or donor-sites, a start lost, or a stop gained or lost) and were present in with low allele frequencies ($f < 0.01$ in all subjects). We tested in two ways: (1) assessing if rare variants within a gene are increased in cases compared to controls and then (2) assessing if rare variants are increased in controls compared to cases. We used the exact statistical framework described above to test for aggregate effects.

Defining CFH Haplotypes—We used sequence-based genotype data from individuals that had 0% missing genotypes for CFH markers used in our previous study to define haplotypes[5], and constructed haplotypes with Beagle[55]. We selected haplotypes with frequencies >0.5%, and calculated case and control frequencies. For each haplotype we calculated odds ratios and 95% confidence intervals relative to the most frequent haplotype.

Adjusting for stratification—We reassessed association statistics for *C3* K155Q and rare *CFI* coding variants after adjusting for stratification in a subset of sequenced samples with exome-chip data (1,558 cases and 683 controls). We applied EIGENSTRAT to calculate principal components to capture ancestry information and exclude outliers in sequenced samples[54]. For sequenced samples we plotted carriers of *C3* K155Q and also of *CFI* rare variants along the first two components to look for gross evidence of stratification. We also reassessed two-tailed association statistics by adjusting for the top 5 principal components in a logistic regression framework. Since logistic regression can be unstable for rare events, in the case of *C3* K155Q we reported two-tailed significance by calculating beta with the actual data, and comparing it to a null beta distribution defined by permuting case-control labels 1,000,000 times.

Adjusting for age and gender—For samples used in the sequencing study, we had information on gender for all samples, and information on age for almost all (2,457) samples. Here age is defined for controls as the age of ophthalmologic evaluation and for cases as the age at which they had their first ophthalmologic evaluation on record confirming advanced disease. We calculated two-tailed association statistics by adjusting for both age and gender with a logistic regression framework. Similar to the approach we took for adjusting for population stratification, for *C3* K155Q we reported two-tailed significance by calculating beta with the actual data, and comparing it to a null beta distribution defined by permuting case-control labels 1,000,000 times.

C3 K155Q functional assays

Reagents—We obtained purified Factor H (CFH) and Factor I (CFI) (Complement Technologies, Tyler, TX); chicken anti-human C3 antibody (Biodesign International, Saco, ME); goat anti-human C3 antibody (Complement Technologies, Tyler, TX); donkey anti-chicken horseradish peroxidase (HRP) linked IgG (Jackson Immunoresearch, Westgrove, PA); rabbit anti-goat HRP linked IgG (Sigma, St. Louis, MO); murine monoclonal anti-human C3d antibody (Quidel, San Diego, CA); and 3,3′,5,5′-Tetramethylbenzidine (TMB) ELISA substrate and SuperSignal Chemiluminescent Substrate (Thermo Scientific, Rockford, IL).

Protein expression—We applied site directed mutagenesis using QuikChange (Agilent Technologies, Santa Clara, CA) to modify wildtype C3 cDNA with the 155K allele to instead contain the 155Q mutation. We produced 155K and 155Q C3 proteins by transient transfection of 293T cells with TransIt 293 (Mirus, Madison, WI) and collected and concentrated cell supernatants three days post-transfection. We treated transfection supernatants with methylamine (MA) to convert native C3 to a C3b-like form, C3 (MA). We quantified C3 by ELISA, surface plasmon resonance and Western blotting as previously described[30].

Ligand binding assays—To assess binding of 155K and 155Q C3 proteins to regulators, we utilized ELISA assays as previously described[30]. Briefly, we coated either soluble membrane cofactor protein (sMCP), CFH or soluble complement receptor 1 (sCR1) (each at 2 µg/ml) on wells in PBS overnight at 4°C followed by an incubation with a blocking buffer at 37°C for 60 minutes. We prepared dilutions of 155K and 155Q C3 proteins in a low salt (25mM) ELISA buffer. Following incubation at 37°C for 1 hour, we washed the wells and then applied affinity purified chicken anti-human C3 Ab (1:10,000) (Biodesign International) for 1 hour at 37°C. After washing, we applied HRP linked donkey anti-chicken IgY (1:10,000) for 1 hour at 37°C. Following washing, we added TMB substrate and quantified absorbance at 630 nm.

Surface plasmon resonance (SPR) analysis—We performed SPR analysis using the BIAcore 2000 (GE Lifesciences, Piscataway, NJ). Using standard amine coupling technology (GE Lifesciences, Piscataway, NJ) sMCP, CFH or anti-C3d mAb were coupled to individual flow paths. We activated one flow path in each chip without protein as a reference. The running buffer was composed of 10 mM Hepes, pH 7.4, 25 mM NaCl and

0.005% Tween-20. We injected 155K or the 155Q C3 protein for 90 sec at 30 μl/min and monitored dissociation for 300 sec. We regenerated the chip by injecting 0.5 M NaCl. We analyzed each protein at four concentrations, with a minimum of two injections per concentration. These experiments were performed three times using independently produced and quantitated C3 preparations. We analyzed data using the BIAeval software from BIAcore.

Cofactor assays—We assessed cleavage of 155K and 155Q C3 proteins by FI using previously described cofactor assays[30]. C3 preparations were incubated for 0 to 30 min at 37°C with CFI (5 ng in MCP and 20 ng in CFH assays) and a cofactor protein MCP (50 ng; recombinant) or CFH (200 ng) in 15 μl of buffer (10 mM Tris, pH 7.4, 150 mM NaCl). To stop the reaction, we added 7 μl of 3X reducing Laemmli sample buffer. The samples were boiled at 95°C for 5 min, electrophoresed on 10% Tris-glycine polyacrylamide gels, transferred to nitrocellulose and blocked overnight with 5% non-fat dry milk in PBS. We probed the blots with a 1:5000 dilution of goat anti-human C3 Ab followed by HRP-conjugated rabbit anti-goat IgG. We developed the blots with SuperSignal substrate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKOWLEDGEMENTS

REFERENCES

1. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. Lancet. 2012; 379:1728–1738. [PubMed: 22559899]

2. Seddon JM, Cote J, Page WF, Aggen SH, Neale MC. The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. Arch Ophthalmol. 2005; 123:321–327. [PubMed: 15767473]

3. Friedman DS, et al. Prevalence of age-related macular degeneration in the United States. Arch Ophthalmol. 2004; 122:564–572. [PubMed: 15078675]

4. Fritsche LG, et al. Seven new loci associated with age-related macular degeneration. Nature Genetics. 2013; 45:433–439. [PubMed: 23455636]

5. Raychaudhuri S, et al. A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. Nature Genetics. 2011; 43:1232–1236. [PubMed: 22019782]

6. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010; 20:1297–1303. [PubMed: 20644199]

7. Goldstein JI, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics. 2012; 28:2543–2545. [PubMed: 22843986]

8. Huyghe JR, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nature Genetics. 2012 doi:10.1038/ng.2507.

9. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am. J. Hum. Genet. 2008; 83:311–321. [PubMed: 18691683]

10. Neale BM, et al. 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. 2010; 107:7395–7400. [PubMed: 20385826]

11. Fagerness JA, et al. Variation near complement factor I is associated with risk of advanced AMD. Eur. J. Hum. Genet. 2009; 17:100–104. [PubMed: 18685559]

12. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nature Methods. 2010; 7:248–249. [PubMed: 20354512]

13. MacArthur DG, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012; 335:823–828. [PubMed: 22344438]

14. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum. Mutat. 2011; 32:661–668. [PubMed: 21480434]

15. Bienaime F, et al. Mutations in components of complement influence the outcome of Factor I-associated atypical hemolytic uremic syndrome. Kidney International. 2009; 77:339–349. [PubMed: 20016463]

16. Caprioli J, et al. Genetics of HUS: the impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome. Blood. 2006; 108:1267–1279. [PubMed: 16621965]

17. Nilsson SC, et al. A mutation in factor I that is associated with atypical hemolytic uremic syndrome does not affect the function of factor I in complement regulation. Mol. Immunol. 2007; 44:1835–1844. [PubMed: 17084897]

18. Geelen J, et al. A missense mutation in factor I (IF) predisposes to atypical haemolytic uraemic syndrome. Pediatr. Nephrol. 2007; 22:371–375. [PubMed: 17106690]

19. Sellier-Leclerc A-L, et al. Differential impact of complement mutations on clinical characteristics in atypical hemolytic uremic syndrome. J. Am. Soc. Nephrol. 2007; 18:2392–2400. [PubMed: 17599974]

20. Fremeaux-Bacchi V, et al. Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome. J. Med. Genet. 2004; 41:e84. [PubMed: 15173250]

21. Kavanagh D, et al. Characterization of mutations in complement factor I (CFI) associated with hemolytic uremic syndrome. Mol. Immunol. 2008; 45:95–105. [PubMed: 17597211]

22. Li M, et al. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. Nature Genetics. 2006; 38:1049–1054. [PubMed: 16936733]

23. Raychaudhuri S, et al. Associations of CFHR1-CFHR3 deletion and a CFH SNP to age-related macular degeneration are not independent. Nature Genetics. 2010; 42:553–5. author reply 555–6. [PubMed: 20581873]

24. Maller J, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nature Genetics. 2006; 38:1055–1059. [PubMed: 16936732]

25. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics. 2003; 33:177–182. [PubMed: 12524541]

26. Maller JB, et al. Variation in complement factor 3 is associated with risk of age-related macular degeneration. Nature Genetics. 2007; 39:1200–1201. [PubMed: 17767156]

27. Yates JRW, et al. Complement C3 variant and the risk of age-related macular degeneration. N. Engl. J. Med. 2007; 357:553–561. [PubMed: 17634448]

28. Nishiguchi KM, et al. C9-R95X polymorphism in patients with neovascular age-related macular degeneration. Invest. Ophthalmol. Vis. Sci. 2012; 53:508–512. [PubMed: 22190594]

29. Wu J, et al. Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. Nat. Immunol. 2009; 10:728–733. [PubMed: 19503104]

30. Fremeaux-Bacchi V, et al. Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome. Blood. 2008; 112:4948–4952. [PubMed: 18796626]

31. Sánchez-Corral P, et al. Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. Am. J. Hum. Genet. 2002; 71:1285–1295. [PubMed: 12424708]

32. Józsi M, et al. Factor H and atypical hemolytic uremic syndrome: mutations in the C-terminus cause structural changes and defective recognition functions. J. Am. Soc. Nephrol. 2006; 17:170–177. [PubMed: 16338962]

33. Manuelian T, et al. Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. J. Clin. Invest. 2003; 111:1181–1190. [PubMed: 12697737]

34. Montes T, Tortajada A, Morgan BP, Rodriguez de Cordoba S, Harris CL. Functional basis of protection against age-related macular degeneration conferred by a common polymorphism in complement factor B. Proc. Natl. Acad. Sci. U.S.A. 2009; 106:4366–4371. [PubMed: 19255449]

35. Heurich M, et al. Common polymorphisms in C3, factor B, and factor H collaborate to determine systemic complement activity and disease risk. Proc. Natl. Acad. Sci. U.S.A. 2011; 108:8761–8766. [PubMed: 21555552]

36. Clark SJ, et al. Tissue-specific host recognition by complement factor h is mediated by differential activities of its glycosaminoglycan-binding regions. J. Immunol. 2013; 190:2049–2057. [PubMed: 23365078]

37. Yu Y, et al. Common variants near FRK/COL10A1 and VEGFA are associated with advanced age-related macular degeneration. Hum. Mol. Genet. 2011; 20:3699–3709. [PubMed: 21665990]

38. Chen W, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc. Natl. Acad. Sci. U.S.A. 2010; 107:7401–7406. [PubMed: 20385819]

39. Seddon JM, Sharma S, Adelman RA. Evaluation of the clinical age-related maculopathy staging system. Ophthalmology. 2006; 113:260–266. [PubMed: 16458093]

40. Seddon JM, Cote J, Davis N, Rosner B. Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio. Arch Ophthalmol. 2003; 121:785–792. [PubMed: 12796248]

41. Seddon JM, Santangelo SL, Book K, Chong S, Cote J. A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. The American Journal of Human Genetics. 2003; 73:780–790. [PubMed: 12945014]

42. Seddon JM, et al. Dietary fat and risk for advanced age-related macular degeneration. Arch Ophthalmol. 2001; 119:1191–1199. [PubMed: 11483088]

43. Yang Z, et al. Genetic and functional dissection of HTRA1 and LOC387715 in age-related macular degeneration. PLoS Genet. 2010; 6:e1000836. [PubMed: 20140183]

44. Yang Z, et al. Toll-like receptor 3 and geographic atrophy in age-related macular degeneration. N. Engl. J. Med. 2008; 359:1456–1463. [PubMed: 18753640]

45. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]

46. Daly AK, et al. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nature Genetics. 2009; 41:816–819. [PubMed: 19483685]

47. Sklar P, et al. Whole-genome association study of bipolar disorder. Mol. Psychiatry. 2008; 13:558–569. [PubMed: 18317468]

48. Rivas MA, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nature Genetics. 2011; 43:1066–1073. [PubMed: 21983784]

49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

50. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011; 43:491–498. [PubMed: 21478889]

51. Tennessen JA, et al. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science. 2012; 337:64–69. [PubMed: 22604720]

52. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6:80–92. [PubMed: 22728672]

53. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007; 81:559–575. [PubMed: 17701901]

54. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006; 38:904–909. [PubMed: 16862161]

55. Browning SR. Multilocus association mapping using variable-length Markov chains. Am. J. Hum. Genet. 2006; 78:903–913. [PubMed: 16685642]
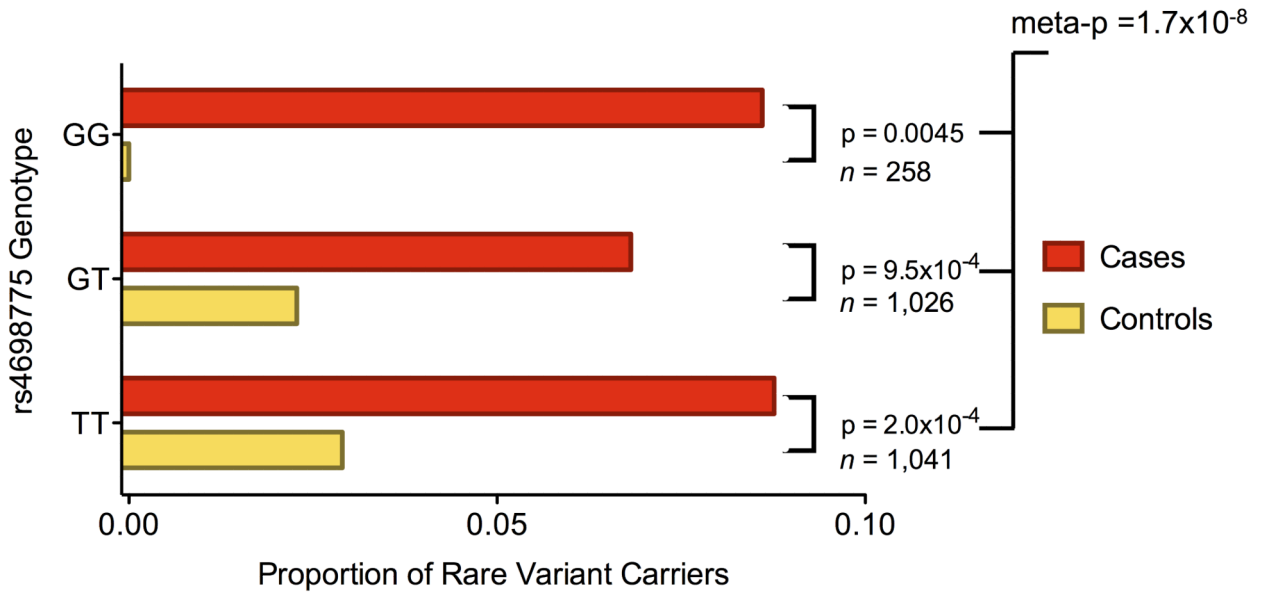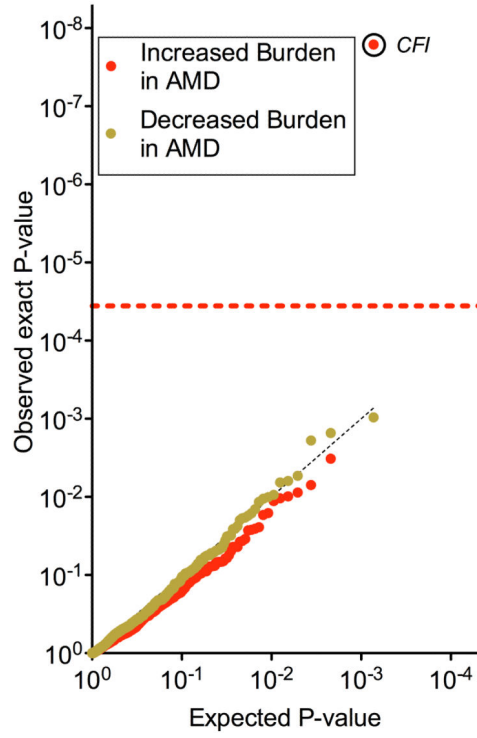
Figure 1. *CFI* burden of rare coding variants is increased in cases
A. We tested 681 sequenced genes for increased burden of rare (<1%) missense, nonsense, splice, or read-through variants. Here we plot the observed *p*-value for each gene as a function of expected *p*-values. We tested enrichment in rare variant carriers in cases (red) separately from enrichment in rare variant carriers in controls (yellow). We indicate the threshold of significance after multiple hypothesis testing (dashed red line, $p < 3.6 \times 10^{-5}$ =
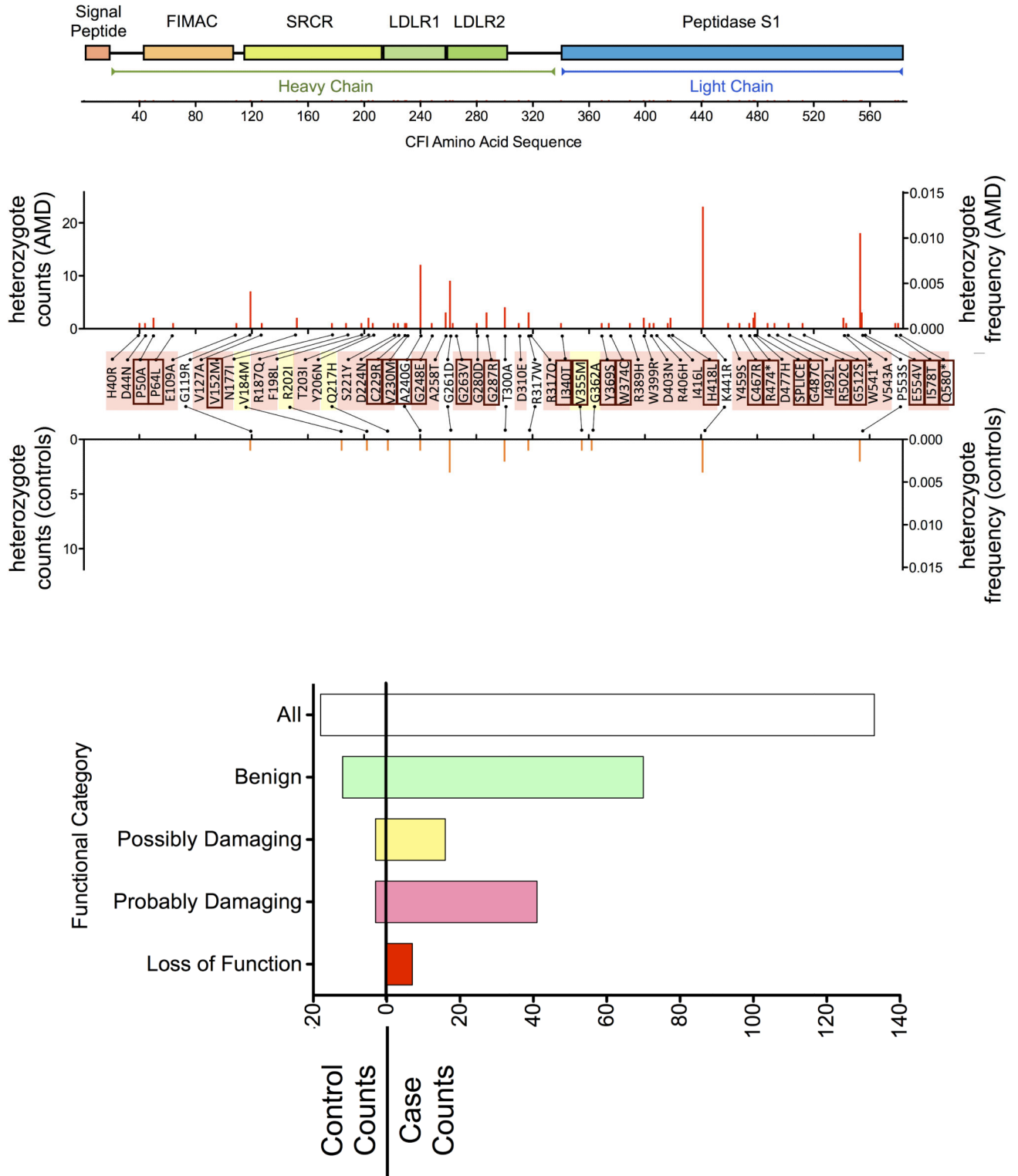
0.05/(681 genes × 2 tests)). We observed significant enrichment of rare variants in cases in only one gene, *CFI*. **B.** We stratified individuals by their rs4698775 genotype, a nearby associated common variant; the G minor allele has previously been reported to be associated with risk. We observed that rare variants are enriched on all genotypic backgrounds.

Figure 2. Burden of CFI rare variants

A. In the top we show the domain structure of the CFI protein, including the factor I membrane attack complex (FIMAC), scavenger receptor cysteine rich (SRCR) domain, the

LDL-receptor (LDLR1 and LDLR2) domains, and a peptidase S1 serine protease domain. Below we plot the individual rare variants that we observed in our sequencing experiment in cases (red, upper) and controls (yellow, lower). In our data, we observed no common *CFI* coding variants with $f>0.01$. We plot counts of variants along the left y-axes, and the proportions of heterozygote individuals along the right y-axes. We list the actual coding changes conferred by the variants in the middle. Case-only variants are colored red and control-only variants are colored yellow. We put red boxes around variants that are predicted to be loss of function or probably damaging variants. **B.** Plot of rare variants in *CFI* categorized with PolyPhen-2 as all variants, "benign", "possibly damaging", "probably damaging", and "loss of function", with counts in controls (left) compared to counts in cases (right) for each category. Variants that are predicted to have greater liklihood of being damaging are more strongly skewed toward cases (1-tailed *p*=0.045, logistic regression with permutation of only individuals with a rare variant)
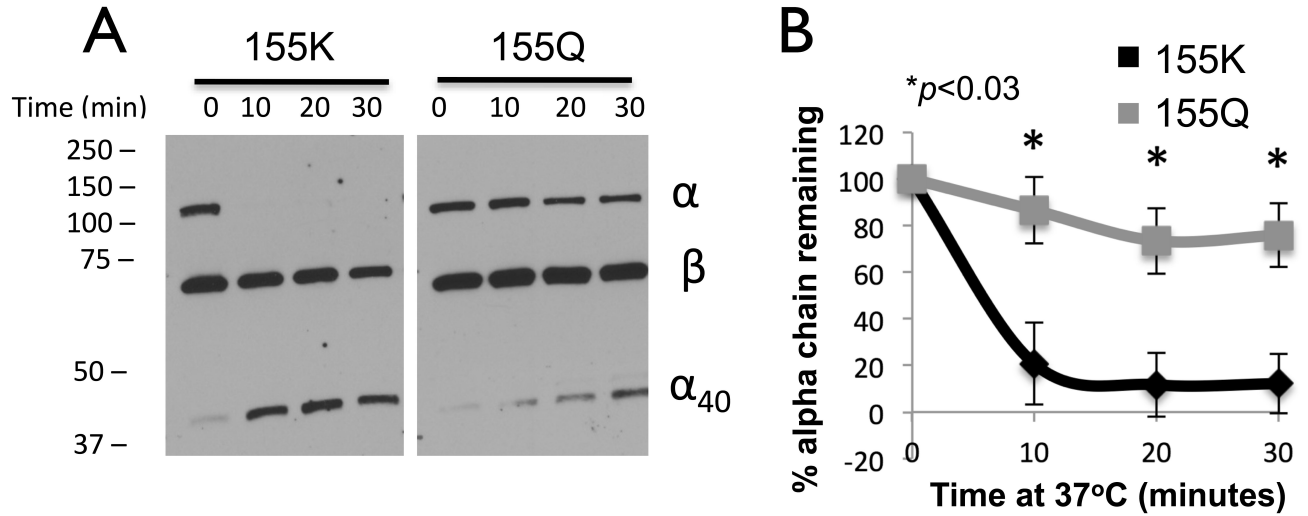
Figure 3. The C3 155Q confers resistance to cofactor activity

A. We incubated 155K (WT) and 155Q (variant) C3 proteins at physiologic salt concentration with 20 ng of CFI and 200 ng of CFH at 37°C for 10, 20, and 30 minutes and then stopped the reaction by the addition of 3X reducing sample buffer. Following electrophoresis and transfer to a nitrocellulose membrane, western blots were performed for C3 proteins. This western blot is representative of four similar experiments. We observe that the α-chain is cleaved much less efficiency in 155Q than in 155K. In parallel we see accumulation of the α_{40} cleavage fragment in proportion to α-chain cleavage. The α_1 cleavage fragment is not visualized since it migrates at the same molecular weight as the β-chain. **B.** We quantified the reduction of the α-chain by densitometry of four independent experiments; here, we present the mean ±SEM at each time point. We assessed significance with an unpaired t-test.

TABLE 1

C3 K155Q (rs147859257) association with AMD risk

Cohort	Case/Control					Discordant Sibpairs			K155Q Association	
	K155Q Heterozygotes		K155 Homozygotes		Allelic OR	Concordant Genotypes	K155Q in affected sib	K155Q in unaffected sib	One-tailed p-value	
	Un-affected	AMD	Un-affected	AMD					Un-conditional	Conditioning on rs2230199
Boston (sequencing, discovery)	1	40	744	1636	18.0 (2.5 − 130.9)	36	0	0	4.8×10^{-6}	9.6×10^{-6}
Boston (replication)	19	20	2014	726	2.9 (1.5 − 5.4)	13	0	0	9.3×10^{-4}	3.6×10^{-4}
French	4	18	676	935	3.2 (1.1 − 9.6)	-	-	-	0.018	0.011
Baltimore	3	17	163	467	2.0 (0.6 − 6.7)	-	-	-	0.24	-
All Replication	26	55	2853	2128	2.8 (1.7 − 4.6)	13	0	0	3.5×10^{-5}	9.2×10^{-6}
Joint	27	95	3597	3764	3.8 (2.3 − 6.1)	49	0	0	5.2×10^{-9}	6.2×10^{-10}

Each row represents association statistics for the sample collections. The first row (Boston, sequencing) represents our discovery data set, and lists genotype counts in the case-control samples and 49 discordant sibling pairs, and the exact one-tailed association p-values. We list one-tailed p-values with and without conditioning for the rs2230199 *C3* R102G common variant. Other rows list similar statistics for each of the replication cohorts. The last row represents the joint analysis of all samples. Since we tested a total of 1,824 variants for risk and for protection initially, the multiple-hypothesis corrected significance threshold for joint analysis is $p < 1.4 \times 10^{-5}$.

Table 2

C9 P167S (rs34882957) association with AMD risk

Cohort	Case/Control					Discordant Sibpairs			P167S Association One-tailed *p*-value
	P167S Heterozygotes		P167S Homozygotes		Allelic OR	Concordant Genotypes	P167S in affected sib	P167S in unaffected sib	Un-conditional
	Un-affected	AMD	Un-affected	AMD					
Boston (sequencing, discovery)	11	53	734	1623	2.2 (1.1 – 4.1)	35	1	0	0.0068
Boston (replication)	39	29	1994	717	2.0 (1.3 – 3.3)	12	1	0	0.0023
French	9	33	671	919	2.6 (1.3 – 5.6)	-	-	-	0.0045
Baltimore	4	21	127	438	1.7 (0.6 – 5.1)	-	-	-	0.22
All Replication	51	79	2828	2128	2.2 (1.5 – 3.2)	12	0	0	2.4×10^{-5}
Joint	63	136	3546	3697	2.2 (1.6 – 3.0)	47	2	0	6.5×10^{-7}

As in Table 1, each row represents association statistics for the sample collections. The last row represents the joint analysis of all samples. The multiple-hypothesis corrected significance threshold for joint analysis is $p<1.4\times10^{-5}$.