Sequence and expression of complement factor H gene cluster variants and their roles in age-related macular degeneration risk


Published in: Investigative Ophthalmology and Visual Science

Document Version: Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal: Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2016, The Authors
This is an open access article published under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Sequence and Expression of Complement Factor H Gene Cluster Variants and Their Roles in Age-Related Macular Degeneration Risk

Anne E. Hughes,1 Stephen Bridgett,2 Weihua Meng,2 Mingyao Li,3 Christine A. Curcio,4 Dwight Stambolian,5 and Declan T. Bradley2

1Formerly of Centre for Public Health, Queen’s University Belfast, Belfast, United Kingdom
2Centre for Public Health, Queen’s University Belfast, Belfast, United Kingdom
3Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, United States
4Department of Ophthalmology, University of Alabama, Birmingham, Alabama, United States
5Department of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Correspondence: Anne E. Hughes, 169 Ballylesson Road, Belfast BT8 8JU, UK; annehughes169@btinternet.com.
Submitted: November 30, 2015
Accepted: April 10, 2016

PURPOSE. To investigate how potentially functional genetic variants are co-inherited on each of four common complement factor H (CFH) and CFH-related gene haplotypes and to measure expression of these genes in eye and liver tissues.

METHODS. We sequenced the CFH region in four individuals (one homozygote for each of four common CFH region haplotypes) to identify all genetic variants. We studied associations between the haplotypes and AMD phenotypes in 2157 cases and 1150 controls. We examined RNA-seq profiles in macular and peripheral retina and retinal pigment epithelium/choroid/sclera (RCS) from eight eye donors and three liver samples.

RESULTS. The haploptic coinheritance of potentially functional variants (including missense variants, novel splice sites, and the CFHR3-CFH1 deletion) was described for the four common haplotypes. Expression of the short and long CFH transcripts differed markedly between the retina and liver. We found no expression of any of the five CFH-related genes in the retina or RCS, in contrast to the liver, which is the main source of the circulating proteins.

CONCLUSIONS. We identified all genetic variants on common CFH region haplotypes and described their coinheritance. Understanding their functional effects will be key to developing and stratifying AMD therapies. The small scale of our expression study prevented us from investigating the relationships between CFH region haplotypes and their expression, and it will take time and collaboration to develop epidemiologic-scale studies. However, the striking difference between systemic and ocular expression of complement regulators shown in this study suggests important implications for the development of intraocular and systemic treatments.

Keywords: complement factor H, genetic variation, gene expression, genetic association, haplotype

Ten years have passed since the association between variation in the human complement factor H (CFH) gene and risk of age-related macular degeneration (AMD) was discovered,1-3 yet this knowledge has not yet been translated into preventive measures for the aging population or treatments for patients with AMD. New therapeutic agents that modify the function of the pathway are in development and in trials, and stratified medicine approaches to the use of existing treatments may soon require ophthalmologists to choose treatments according to individual patients’ inherited genetic variants.

Dysfunction of the complement system (a central part of the innate immune system) is key to the etiology of AMD.4 The complement system comprises a series of proteins, C1 to C9, which are sequentially activated or inhibited by regulatory molecules, including complement factor H (FH), factor D (FD), and factor I (FI). The central component of the process, C3, forms enzymes that accelerate its own activation in a positive feedback loop. When in homeostatic balance, the system facilitates destruction of microbial pathogens, apoptosis, and the removal of cell debris without damage to healthy self-tissues. Pharmaceuticals currently in clinical trials that inhibit the complement system include lampalizumab (used to inhibit FD, currently in phase III trials for the treatment of “dry” AMD or geographic atrophy in people with specific genotypes), compstatin (an inhibitor of C3, in phase II trials for treatment of AMD), and bikacimab (an inhibitor of FB that is in phase II trials for treatment of AMD).5 Numerous studies of the impact of patients’ genetics on their response to intravitreal anti-VEGF for AMD have been carried out, yet they have primarily focused on the effect of only one polymorphism, Y402H, even though there are four distinct haplotypes of CFH with different functional variants and risk profiles. Even while potential clinical applications are being investigated there are many unanswered questions about how the complement system behaves in people with different inherited genetic variants and in the different tissues of the eye.
Sequence and Expression of CFH Gene Variants

The major inhibitor of the system, FH, is produced in the liver and secreted into the serum, where it is highly abundant. Rare CFH mutations are associated with early-onset AMD. Relatively little is known about how variants influence pathogenesis. Factor H protein is composed of 20 short consensus repeats (SCRs), each approximately 60 amino acids in length, which share homology at specific residues. Full-length FH is encoded by 22 exons, and a shorter form, FHL-1, stops after splicing of an alternative 10th exon. Five homologous CFH-related genes (CFHR1 to CFHR5 encoding FHR-1 to FHR-5) exist in tandem on chromosome 1q in a region of strong linkage disequilibrium (Supplementary Fig. S1). Four common haplotypes of the CFH region are inherited in white Europeans, and each confers a different risk of AMD (and other diseases). A common deletion of entire genomic region. Sequencing was performed on a HiSeq2000 (Illumina, San Diego, CA, USA) with 100-bp paired-end reads. We aligned genomic reads to hg19 human reference sequence with the Burroughs-Wheeler Aligner (BWA) 0.5.9 aln algorithm and used SAMtools 0.1.14\(^\text{18}\) for sorting, indexing, and removal of duplicate reads. We employed Genome Analysis Toolkit (GATK) to recalibrate and realign and to call polymorphisms. We predicted the effect of coding polymorphisms with PolyPhen-2 (Polymorphism Phenotyping v2).

Investigation of the Effect of CFH Region Haplotypes on AMD Phenotypes

To investigate the effect of polymorphisms on the differential risk on effect of neovascular AMD compared to drusen only (as a proxy for progression), we conducted a genome-wide case-control study and candidate gene studies using the Michigan, Mayo (AREDS, Pennsylvania (MMAP; dbGAP accession 000062; v1.p1)) AMD study cohort consisting of 2157 cases of neovascular AMD, geographic atrophy, or drusen and 1150 healthy controls of white European ancestry (Supplementary Table S1). The candidate gene studies aimed to investigate the roles of the four CFH haplotypes and a representative polymorphism from each of the other known major AMD loci (CFB, C3, and ITIH4\(^\text{1}-\text{3,22-25}\)) on the differential risk of neovascular AMD compared to drusen. The risks of drusen and neovascular AMD compared to disease-free controls are shown to illustrate the effects associated with each haplotype.

Materials and Methods

Study Populations

Institutional Review Board (IRB)/Ethics Committee approval was obtained, and informed consent was obtained from all participants in the original studies. The research adhered to the tenets of the Declaration of Helsinki. DNA for the sequencing experiment was from the anonymized DNA, RNA, and Serum Bank in the Centre for Public Health at Queen’s University Belfast. Patients had neovascular AMD confirmed by clinical examination, grading of digital fundal color photographs, and fluorescein angiography. No details about the clinical course or treatment were available. Research ethics approval for this bank allows analysis of DNA samples for this purpose (Office for Research Ethics Committees Northern Ireland approval reference 11/NI/0139). The in silico data analyzed for this report were from IRB-approved studies. The characteristics of participants included in these studies are summarized in Supplementary Table S1.

Investigation of CFH and CFH-Related Gene Sequence

To identify functional variants in the CFH gene region, we selected genomic DNA samples from individuals of white European ancestry who were homozygous for each of the four common CFH haplotypes, as described previously. Full details of the preparatory, sequencing, and bioinformatics methods have been provided (Supplementary Methods). Briefly, we quantified and fragmented genomic DNA, size selected, and indexed it for Illumina sequencing using the TrueSeq system. We used Nimblegen SeqCap EZ capture to target the entire CFH and CFHR genomic region. Sequencing was performed on a HiSeq2000 (Illumina, San Diego, CA, USA) with 100-bp paired-end reads. We aligned genomic reads to hg19 human reference sequence with the Burroughs-Wheeler Aligner (BWA) 0.5.9 aln algorithm and used SAMtools 0.1.14\(^\text{18}\) for sorting, indexing, and removal of duplicate reads. We employed Genome Analysis Toolkit (GATK) to recalibrate and realign and to call polymorphisms. We predicted the effect of coding polymorphisms with PolyPhen-2 (Polymorphism Phenotyping v2).

Investigation of FH, FHL-1, and FHR-1 to FHR-5 Expression in the Retinal Pigment Epithelium, Choroid, and Sclera and in the Retina

To investigate the relative expression of FH, FHL-1, and FHR-1 to FHR-5 in eye tissues and to investigate the expression of alternatively sliced variants, we aligned RNA-seq reads from our (ML, CAC, DS) previous studies\(^\text{29}\) to the human reference hg19 genome using GSNAP. Eight participants of white European ancestry from the United States were included. Full details are described in Supplementary Methods. We used SAMtools (version 0.1.19)\(^\text{30}\) to sort reads; Picard (version 1.121)\(^\text{30}\) to mark duplicate reads; and SAMtools\(^\text{18}\) depth to obtain read depth. Relative read depths were calculated for FH and FHL-1 relative to FH and FHL-1 combined. An estimate of the proportion of FH relative to both FH and FHL-1 combined for...
all samples was calculated by random effects meta-analysis using metaprop in meta (v4.0-3) in R 3.2.1. This method was chosen as it allowed calculation of an appropriately weighted average proportion (taking into account the total number of reads from each individual) with an overall estimate of uncertainty (confidence interval [CI]) for each tissue.

**Investigation of FH, FHL-1, and FHR-1 to FHR-5 Expression in Liver**

To investigate the relative expression of FH, FHL-1, and FHR-1 to FHR-5 in liver, we accessed liver RNA-seq reads from three individuals (no details of ancestry were available) from the EBI Illumina body map and EBI Expression Atlas and aligned them to the human reference hg19 chromosome 1 using STAR aligner (version 2.4.0f1). We sorted the alignments with the human reference using STAR uncertainty (confidence interval [CI]) for each tissue.

**RESULTS**

**Sequencing of CFH and CFH-Related Genes From Genomic DNA**

We identified nine missense coding SNPs (cSNPs) within CFH and the CFH-related genes, each of which had a minor allele that was homozygous and restricted to one of the four common haplotypes (Table 1). In addition, four heterozygous missense cSNPs (interpreted as rare SNPs against specific haplotype backgrounds; Table 1) and numerous synonymous cSNPs were also identified (Supplementary Table S2). Haplotype A carried a SNP that was predicted to create a novel splice site. An ancestral gene conversion event has caused the replacement of part of the CFHRI gene with part of FH on haplotype C only, resulting in the FHR-1 157Y, 159V, and 175Q polymorphisms in SCR 3. This means that on haplotype C, SCR 3 of FHR-1 is identical to SCR 3 of FH, which created a premature splice donor site GCAAG>AAGG, resulting in replacement of residues 140 to 144 by a single cysteine residue, removing one of the key cysteine residues responsible for the SCR structure.

**Investigation of Risk of AMD Phenotypes**

The risks of drusen and neovascular AMD associated with each CFH haplotype were calculated to illustrate the effects of each CFH haplotype (Table 2). We compared genome-wide variation in individuals with neovascular AMD to those with drusen and found that two SNPs were significantly associated with differential risk (a proxy for a “progression” phenotype) after Bonferroni correction for multiple testing: rs932275 and rs2242799, both in HTRA1 (Supplementary Table S5). We identified several SNPs with a suggestive P value < 5x10^-8, many of which were in biologically plausible genes (Supplementary Table S3). We also compared individuals with neovascular AMD to those with drusen for previously reported major AMD loci (Supplementary Table S4), but the only association identified from among these candidates was the same one that was identified at the HTRA1 locus in the genome-wide analysis. The complement system gene variants were not significantly different between people with drusen and people with neovascular AMD in this cohort.

**Retinal Expression of CFH and CFH-Related Genes**

We found no expression of any CFH-related genes (CFHR1-5) in any of the eye tissues studied. Expression of both long and short CFH transcripts was 50- to 50-fold higher in the RCS than in peripheral or macular retina (Table 3). Within the retina, the full-length FH transcripts were considerably more abundant than FHL-1 transcripts in the macula compared to the periphery (Table 3). The RCS showed increased ratio of FH to FHL-1 transcripts in the periphery compared to the macula in seven of eight eyes (Supplementary Table S5).

**Expression of CFH and CFH-Related Genes in Liver and Confirmation of the Effect of rs4085749 T on CFHRI mRNA**

Relative to total CFH transcription in the liver, CFHRI was expressed approximately 2-fold more strongly, which created a premature splice donor site GCAAG>AAGG, resulting in replacement of residues 140 to 144 by a single cysteine residue, removing one of the key cysteine residues responsible for the SCR structure.
similar level, and CFHR3 approximately 24-fold less strongly (Supplementary Table S6). CFHR4 and CFHR5 showed lower expression, between 5% and 17% of CFH levels. Relative expression of CFH long (FH) and short (FHL-1) transcripts varied widely, with reduction of FH transcripts in proportion to the number of copies haplotype C carried (Supplementary Table S7).

CFHR2 reads from the haplotype AC heterozygote showed that the early splice donor site created by the T allele of rs4085749 carried on haplotype C was used for 66% of exonspansing reads, while the original site 12 bases downstream was used for the remaining 34% of reads from this haplotype. Reads from the homozygous haplotype CC sample showed 68% reduction of read depth across the site, and 60% of exonspansing reads used the novel splice site (Supplementary Fig. S3).

The liver RNA-seq reads also revealed several novel exons that were occasionally found in other transcripts. Of note was an exon within intron 4 of CFHR3 (196,757,875–196,758,074) that was included in about half of all CFHR3 transcripts, despite carrying an in-frame stop codon.

**DISCUSSION**

The central role of factor H in innate immunity and the effects of its genetic variation on risk of AMD, atypical hemolytic uremic syndrome (aHUS),9 systemic lupus erythematosus,36 rheumatoid arthritis,50 and invasive meningococcal infection9 make it intriguing, and it has yielded many surprises. Our study is the first to investigate the extended CFH region in individuals who are homozygous for the common European haplotypes. The results of this study should provide a clear foundation for future studies of FH function and aid the development of personalized approaches based on patients’ own complement system genetics.

Haplotype A confers greatest risk of AMD and carries the 402H risk variant.12 Codon 402 is found in both FH and FHL-1, so both products of the CFH gene are affected.40 This haplotype also carries a possible new splice site (rs203685), which could reduce expression of CFH through nonsense-mediated decay. Haplotype A produces the acidic isoform of FHR-1.

Haplotype B has no effect on AMD risk,12 but confers reduced risk of meningococcal disease, which may be due to altered binding of meningococci to FH or FH-related proteins.9 We showed that the minor alleles of virtually all variants within the CFHR3 region were found on haplotype B. CFH 936D and CFHR3 2418, two variants in the 3’ untranslated region (UTR) of CFHR3 (rs402372 and rs390837), and CFHR2 72Y are carried on haplotype B. This haplotype also produces the acidic isoform of FHR-1.

Haplotype C protects against AMD.12 CFHR1 157Y, 159V, and 175Q, which have arisen by conversion from CFH, code for the basic isoform of FHR-1 on this haplotype. This may result in altered competition between FHR-1 and FH for binding sites, or alteration of the regulatory roles of FHR-1. On haplotype C, CFH 62I encodes FH with increased binding activity for C3b, and increased cofactor activity for complement factor 1-mediated cleavage of fluid phase and cell-bound C3b.41 The preferential use of an alternative splice donor site on this haplotype may have functional importance.42 This site is predicted to result in omission of one of the four cysteine residues that form the SCR structure and would have a major effect on the protein structure. Haplotype D is missing the entire CFHR3 and CFHR1 genes.14 There are no cSNPs in CFH on this haplotype. Homozygotes for haplotypes C and D are at elevated risk of aHUS.8,15,43,44 Absence of FHR-1 is thought to be critical, resulting in lack of immune tolerance to an epitope of FHR-1 that mimics a neoepitope of FH that may be formed during interaction with bacteria and development of antibodies that cross-react against FH.15,45

Although CFH was expressed in the retina and in the RCS of normal aged eyes, we found no evidence of transcription of any of the CFHR-related genes (CFHR1-5) in these tissues. This finding contrasts with reports by Bennis et al.45 and Booij et al.46,47 of a microarray experiment that measured transcriptome expression of six human eyes, which reported CFHR1 transcription at approximately one-fourth the level of CFH (and lower levels of CFHR4, CFHR5, and CFHR2, but not CFHR3). They found no CFHR1 expression in mouse eye tissues, which is consistent with an earlier report by Luo et al.48 in which RT-PCR indicated no expression of CFHR1 in mouse retina or RPE/chorioid tissue. The conflicting results indicate either that our method did not detect CFHR mRNA where it was present, or that this locus is not expressed in the retina. Alternatively, the Bennis et al.45 and Booij et al.46,47 experiments falsely detected CFHR mRNA. If CFHR genes are expressed in these tissues, the most likely explanation for their apparent absence in our experiment would be differences in the tissue processing between the two studies. Alternatively, the Bennis et al.45 and Booij et al.46,47 custom microarrays may have
suffered cross-reactivity from FH and FHL-1 transcripts (Supplementary Fig. S1), leading to the apparent presence of CFHR genes. Absent expression would not necessarily indicate that CFHR genes have no role in these tissues, as it is likely that they could diffuse across Bruch’s membrane, as FHL-1 does (though FH does not).49 Further studies are needed to assess CFHR expression (if any) in AMD eyes and the effect of inflammation. CFH transcripts for FH and FHL-1 showed relatively low expression in the macular and peripheral retina and much greater expression in the RCS, suggesting reduced protection in the retina and greater protection in at least one of the RCS tissues. Our results suggested a greater proportion of FH compared to FHL-1 in the macular retina compared to the peripheral retina. This may be due to regulation by microRNAs. Both transcripts appear to be heavily regulated by microRNAs; however, target sites for miR146a and miR155 are adjacent on the 3′ UTR of CFH and absent from the 3′ UTR of the short (FHL-1) transcript. These are among the microRNAs with greatest influence on immune signalling pathways and immune homeostasis.50

In contrast to the eye, we found that all CFH-related transcripts were present in liver RNA. Both CFH transcripts, CFHR1, CFHR2, and CFHR3 were much more highly expressed than CFHR4 or CFHR5. Our analysis of liver RNA-seq data suggested that haplotype C may be associated with either considerable reduction in FH or increase in FHL-1 mRNA transcript levels compared to the other haplotypes. Unexpectedly, there seems to be poor correlation between levels of transcripts in the liver and the reported concentrations of their translated products in the blood plasma. FHL-1 is reported at only 2% to 8% of FH levels in plasma, despite having transcript levels in the liver greater than that of FH mRNA.9 It may be important to understand the stability of FH, FHL-1, and their degradation products. Measurement of FH and related proteins is difficult because of the potential for related proteins to cross-react with antibodies used in immunoassays. Reported plasma concentrations of FH also include FHL-1, which reacts with antibodies specific for the N-terminal SCR5 common to both proteins, and there is also considerable scope for cross-reactivity with FH-related proteins. Scholl et al.51 reported increased concentrations of most complement proteins and activation products in AMD compared to normal controls, compatible with systemic complement activation and low-grade inflammation. Complement activation products were highest in those carrying CFH risk haplotypes for AMD.51 In a study that used an immunoassay with antibodies highly specific for either FH 402H or 402Y, Hakobyan et al.52 showed that plasma FH levels of normal adults were not significantly different in those with FH 402Y, 402YH, and 402HH genotypes. The largest study of genetic influences on plasma FH and AMD was reported by Ansari et al. in 2013.53 They showed a relationship between CFH genotypes and concentrations of FH and of FHR-1, though they reported plasma FHR-1 concentrations in people with no copies of the CFHR1 gene to be >50% of those with two copies of the gene, suggesting that their assay cross-reacted with other proteins.53 It is possible, though difficult, to create antibodies against specific variants, and gene conversion adds another layer of complexity to the measurement and assessment of function in this region.

We conducted a secondary analysis of a genome-wide association study of AMD21 to investigate whether variation was associated with progression from drusen to neovascular AMD. Only SNPs in HTRA1 (near ARMS2) showed an association when comparing individuals with neovascular AMD to those with drusen. Variants in CFH, C3, and CFB increased risk of drusen and AMD equally, but added no extra risk effect to differentiate those with drusen and those with neovascular AMD, suggesting the possibility that complement dysfunction is necessary but not sufficient for progression to neovascular AMD. Interestingly, earlier case-control studies reported similar findings for CFH and ARMS2/HTRA1.54,55 whereas a longitudinal study showed a role for CFH on progression from early to late disease over a period of 6 years.56 It is possible that in the case-control studies, age, sex, smoking, and/or other factors might confound or interact with genetic risk factors in a manner not seen in the cohort study.

The most important limitation of our study is that the number of donated eyes is small, and availability of a greater number of eye samples would improve our ability to draw firm conclusions about the levels of FH and FHL-1 expression in people with different CFH haplotypes. Collaborative studies may be required in the future to bring together enough expression data to truly explore the effects of haplotypes on the levels of expression in different eye tissues.

We have comprehensively characterized the variants that may be the underlying causes of the AMD risk conferred by the four common haplotypes. Taking into account the real-world biological complexity of the region may allow some clearer thought about the mechanisms of disease and protection, as well as the potential for future development of personalized treatments based on these mechanisms.

**Acknowledgments**

We thank the participants in the National Eye Institute Study of Age-Related Macular Degeneration (NEI-AMD) and the NEI-AMD Research Group for their valuable contribution to this research. The AMD dataset used for the analyses described in this paper was obtained from the NEI-AMD database, through dbGaP accession no. phs000182.v1.p1. Funding support for NEI-AMD was provided by the National Eye Institute.

Supported by Guide Dogs for the Blind Association Grant OR2010-03c (AEH), National Eye Institute Grant R01EY023164, and the Arnold and Mabel Beckman Initiative for Macular Research (CAC and DS).

Disclosure: A.E. Hughes, None; S. Bridgett, None; W. Meng, None; M. Li, None; C.A. Curcio, Genentech (C), Novartis (C), Merck (C), Janssen Cell Therapy (C), IOVS (S); D. Stambolian, None; D.T. Bradley, None.

**References**


Sequence and Expression of CFH Gene Variants


