# Haplotypes Spanning the Complement Factor H Gene Are Protective against Age-Related Macular Degeneration

*Kylee L. Spencer,*[1] *Michael A. Hauser,*[2] *Lana M. Olson,*[1] *Nathalie Schnetz-Boutaud,*[1] *William K. Scott,*[3] *Silke Schmidt,*[2] *Paul Gallins,*[4] *Anita Agarwal,*[1] *Eric A. Postel,*[2] *Margaret A. Pericak-Vance,*[3] *and Jonathan L. Haines*[1]

**Purpose.** Age-related macular degeneration (AMD) is a devastating disorder that adversely affects the quality of life of nearly 2 million Americans who have advanced forms of the disease. Besides the well-known risk imparted by carrying the Y402H variant in the complement factor H (CFH) gene on chromosome 1, recent evidence of the existence of protective haplotypes spanning CFH has been reported.

**Methods.** The haplo.stats program was used to test for association of the protective haplotypes after adjusting for age in the dataset of 584 sporadic cases and 248 control samples. Logistic regression modeling and likelihood ratio tests were used to investigate an interaction between a particular haplotype and smoking status. The HBAT option of FBAT was used to confirm the associations in an independent dataset of 201 families.

**Results.** Two protective (P) haplotypes in a family-based dataset (P1 = CAATTTAG, $P = 0.021$; and P2 = CGGCTTAG, $P = 0.018$) were identified for the first time. Age-adjusted score statistics provided support for these protective haplotypes in the case–control dataset (P1 frequency in cases ~13%, in controls ~20%, $P = 0.001$; P2 frequency in cases ~5%, in controls ~8%, $P = 0.077$). There was also tentative evidence of an interaction between one of the protective haplotypes and cigarette smoking ($P = 0.04$ likelihood ratio test for P2–smoking interaction).

**Conclusions.** Replication of the association between the protective haplotypes and decreased AMD susceptibility provides increased evidence that these associations have biological meaning. The suggestion of a haplotype–smoking interaction adds to the growing body of evidence that smoking is an important environmental covariate in AMD that should be considered in genetic studies. Identification of the protective variant(s) carried within these haplotypes is critical for understanding the etiology of AMD. (*Invest Ophthalmol Vis Sci.* 2007;48:4277–4283) DOI:10.1167/iovs.06-1427

Age-related macular degeneration (AMD) is a complex, late-onset disease that remains one of the primary causes of blindness in the elderly. Large, soft drusen representing undigested cellular debris manifest early in the course of the disease. Later, central visual field loss occurs as geographic atrophy and/or choroidal neovascularization develop.[1] A meta-analysis of population-based data from seven studies showed that an estimated 7.3 million people in the United States have large drusen characteristic of early AMD and that 1.75 million individuals have advanced forms of the disorder.[2]

Both genetic and nongenetic risk factors are known to be involved in the pathogenesis of AMD. Epidemiologic studies have successfully identified many nongenetic or lifestyle risk factors for AMD, including increased age, smoking, and body mass index.[3,4] The connection between smoking and risk of AMD has been particularly well documented. A recent literature review reported that 13 of 17 studies showed a statistically significant association between smoking and AMD.[5] Pooled data from 9523 participants in three population-based cohort studies in the United States, The Netherlands, and Australia show an estimated odds ratio for current smokers versus never smokers at 2.83 (1.15–6.93) for geographic atrophy and 2.35 (1.30–4.27) for late AMD, which includes cases with geographic atrophy and choroidal neovascularization.[6] The adjusted population attributable risk percentage for smoking is 20%, suggesting that smoking is likely to be the largest known modifiable risk factor for AMD.[7]

A strong genetic component has long been suspected in the etiology of AMD, based on the results of familial aggregation studies,[8–10] segregation analysis,[11] and twin studies.[12–17] A meta-analysis of the six largest genomic linkage screens identified the chromosome regions 10q26, 1q, 2p, 3p, and 16 as the most likely regions to harbor AMD susceptibility loci.[18] In Caucasian populations, the association between AMD risk and the Y402H variant in the complement factor H (CFH) gene located on chromosome 1 has been validated by multiple studies,[19–27] and promising initial reports of a strong association between a polymorphism in the *LOC387715* gene on chromosome 10 and AMD risk have also been replicated.[7,24,28,29] A recent meta-analysis estimates that the odds ratio for rs10490924 in *LOC387715* as 2.5 (95% confidence interval [CI], 2.2–2.9) in heterozygotes for the risk allele and 7.3 in homozygotes (95% CI, 5.7–9.4).[30] In addition, genetic and lifestyle risk factors for AMD have been tied together by the discovery that the effect of the *LOC387715* variant on AMD risk is modified by cigarette smoking.[7] The most recently reported major AMD susceptibility loci are the factor B (*BF*) and complement component 2 (*C2*) genes in the major histocompatibility complex class III region of chromosome 6.[29,31]

From the [1]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee; the [2]Center for Human Genetics, Duke University Medical Center, Durham, North Carolina; the [3]Department of Medicine and Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; and the [4]Miami Institute for Human Genomics, University of Miami, Miami, Florida.

Supported by Grants EY12118 (MAP-V, JLH) and EY015216 (SS) from the National Eye Institute, and Grant M01 RR-00095 from the National Center for Research Resources (Vanderbilt University).

Submitted for publication December 1, 2006; revised April 16, 2007; accepted July 11, 2007.

Disclosure: **K.L. Spencer**, None; **M.A. Hauser**, None; **L.M. Olson**, None; **N. Schnetz-Boutaud**, None; **W.K. Scott**, None; **S. Schmidt**, None; **P. Gallins**, None; **A. Agarwal**, None; **E.A. Postel**, None; **M.A. Pericak-Vance**, None; **J.L. Haines**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Jonathan L. Haines, Center for Human Genetics Research, 519 Light Hall, Vanderbilt University Medical Center, Nashville, TN 37232; jonathan@chgr.mc.vanderbilt.edu.

**TABLE 1.** Diagnostic AMD Grading

| Grade | Diagnostic Criteria |
|---|---|
| 1 | No drusen or small nonextensive drusen (less than 63 μm diameter), without RPE pigment abnormalities |
| 2 | Extensive small drusen (total extent greater or equal to an area of a circle 125 μm in diameter) or nonextensive intermediate drusen (drusen greater than 63 μm but not greater than 125 μm in diameter), and/or RPE abnormalities associated with AMD (including focal hypopigmentation, focal hyperpigmentation, pigment clumping, and focal retinal pigment epithelial atrophy not large enough to be considered geographic atrophy |
| 3 | Extensive intermediate drusen or any large soft drusen (greater than 125 μm in size), includes drusenoid retinal pigment epithelial detachments (RPED) |
| 4 | Geographic atrophy (area of RPE atrophy with sharp margins, usually with visible choroidal vessels, minimum diameter 300 μm), with or without involvement of the center of the macula |
| 5 | Serous (nondrusenoid) or hemorrhagic retinal pigment epithelial detachments, choroidal neovascularization, subretinal or sub-RPE hemorrhage or fibrosis, or photocoagulation scar consistent with treatment of AMD |

The precise role of CFH in the AMD process is still being elucidated, but it is thought that the nonsynonymous substitution of histidine for tyrosine in exon 9, which lies in the protein's heparin binding site, may disrupt the ability of CFH to attenuate complement response and lead to damage of ocular tissue.[22] Recently, data supporting the existence not only of a risk haplotype, but also of protective haplotypes in the CFH gene have been reported.[19,32,33] Hageman et al.[19] used a tag single nucleotide polymorphism (SNP) approach to select eight SNPs in the CFH gene for haplotype analysis in a large case–control sample of individuals of European-American descent ascertained at the University of Iowa (Iowa City, IA) and Columbia University (New York, NY). Two haplotypes had protective effects, with the SNPs that distinguish the protective haplotypes concentrated between exons 2 and 11. Using a subset of four of these eight SNPs, Okamoto et al.[32] confirmed a protective haplotype in a Japanese case–control dataset. Whereas this haplotype had a frequency of 12% in cases versus 21% in controls in the individuals of European-American descent and an odds ratio of 0.54 (0.41–0.73) in the Japanese sample, the haplotype had a frequency of 18% in cases and 35% in controls and an odds ratio of 0.63 (0.26–0.67).

In contrast, CFH SNPs in a separate Caucasian dataset have also been selected for haplotype analysis using a stepwise regression approach.[33] Notably, the Y402H variant was not selected as having the strongest influence on AMD risk. Instead, five SNPs spanning the latter half of the CFH gene from intron 9 to 18 were chosen. Two of the four most common haplotypes had protective effects (haplotype 1 frequency of 14% cases, 44% in controls; haplotype 2 frequency of 3% in cases, 11% in controls).

Our goal was to study the presence of the risk and protective haplotypes in CFH and to examine the possibility of a haplotype–smoking interaction, as smoking appears to be one of the most important environmental contributors to AMD, and previous studies have shown that the effect of genetic risk factors for AMD may be modified by smoking.[7]

## MATERIALS AND METHODS

### Study Participant Ascertainment

Multiplex families and an independent dataset of unrelated cases and controls, all of Caucasian, non-Hispanic descent, were ascertained at Vanderbilt University Medical Center and Duke University Medical Center. All patients and controls underwent an eye examination and had stereoscopic fundus photographs graded according to a modified version of the Age-Related Eye Disease Study (AREDS) grading system, as described elsewhere.[34,35] Briefly, grades 1 and 2 represent controls. Grade 1 controls have no evidence of drusen or small nonextensive drusen without pigmentary abnormalities, whereas grade 2 controls may show signs of either extensive small drusen or nonextensive intermediate drusen and/or pigmentary abnormalities. Grade 3 AMD cases have extensive intermediate drusen or large, soft drusen with or without drusenoid retinal pigment epithelial detachment. Grade 4 AMD cases exhibit geographic atrophy and grade 5 individuals have exudative AMD, which includes nondrusenoid retinal pigment epithelial detachment, choroidal neovascularization, and subretinal hemorrhage or disciform scarring (Table 1). Individuals were classified according to status in the more severely affected eye. Table 2 describes additional features of the datasets, including age of examination, gender, and a brief description of family structure for the family-based

**TABLE 2.** Study Population Characteristics

| | | Independent Case–Control Dataset | |
|---|---|---|---|
| | | 584 Cases | 248 Controls |
| Total Individuals | Family Dataset 451 (130 Multiplex Families) | (Grades 3, 4, 5) | (Grades 1, 2) |
| Grade (%) | 3 (35.4) | 3 (29.8) | 1 (68.1) |
| | 4 (14.6) | 4 (11.8) | 2 (31.9) |
| | 5 (50.0) | 5 (58.4) | |
| Mean age (SD) | 73.3 (10.3) | 76.5 (7.6) | 66.1 (9.1) |
| Female (%) | 67.6 | 64.0 | 56.5 |
| Ever smoked (%) | 49.5 (unaffected) | 60.4 | 51.0 |
| | 57.8 (affected) | | |

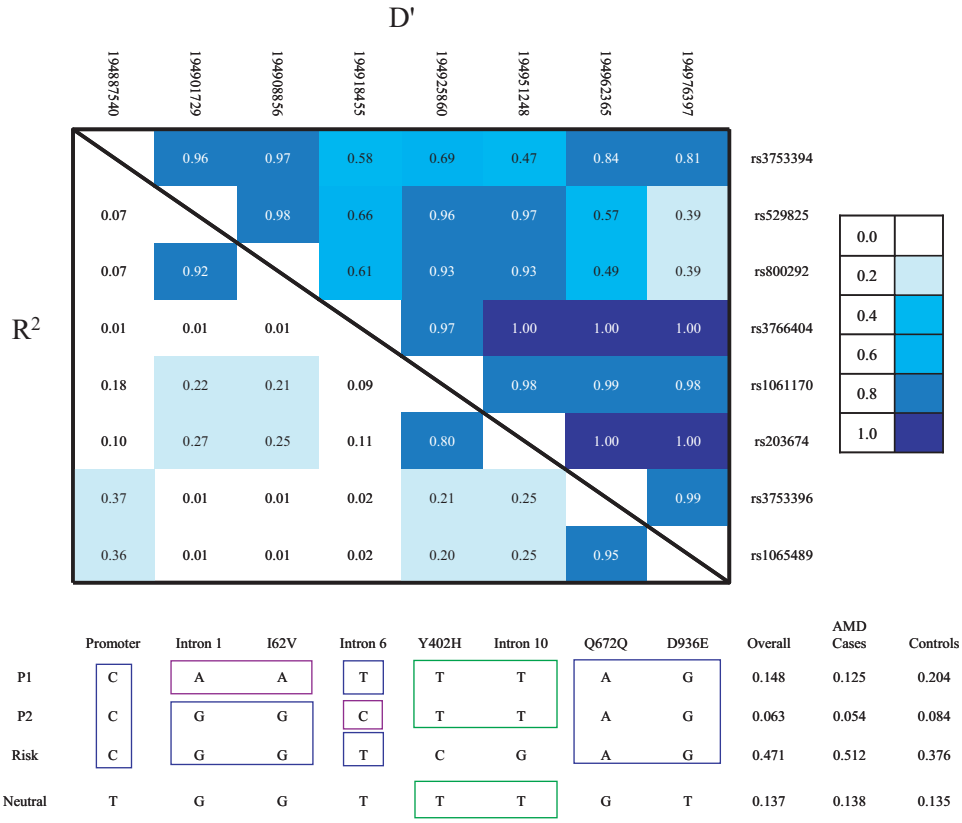


**FIGURE 1.** Linkage disequilibrium and haplotype frequencies in cases and controls. *Blue boxes*: alleles shared between one or both protective haplotypes and the risk haplotype. *Green boxes*: alleles shared between both protective haplotypes and the neutral haplotype. *Purple boxes*: alleles unique to each of the protective haplotypes.

dataset. Approval for the study was obtained from the appropriate institutional review boards at VUMC and DUMC, all study participants gave informed consent, and the research adhered to the tenets of the Declaration of Helsinki.

## DNA Analysis

A commercial system was used to extract genomic DNA from whole blood samples according to standard protocol (Purgene; Gentra Systems, Minneapolis, MN). Genotyping assays (TaqMan Assays-on-Demand or, when necessary, Assays-By-Design; Applied Biosystems, Inc., Foster City, CA) were used for genotyping SNPs in CFH (GenBank accession no. DQ233256 Version DQ233256.1 GI:77744384; http://www.ncbi.nlm.nih.gov/Genbank; provided in the public domain by the National Center for Biotechnology Information, Bethesda, MD), in both the family-based and case–control datasets. The Y402H polymorphism rs1061170 was genotyped by direct sequencing, as previously described.[22] DNA samples from The Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH) families were duplicated between and across plates for use as a quality control, and laboratory personnel were blinded to the affection status of the individuals being genotyped.

## Statistical Methods

Haploview[36] was used to examine the linkage disequilibrium (LD) patterns between SNPs in each dataset, to verify that all SNPs were in Hardy-Weinberg equilibrium, and to perform single marker tests of association. Of the 33 SNPs we genotyped within the region of complement activation (RCA) gene cluster, alleles at eight tagging SNPs in the CFH gene identical with those of Hageman et al.[19] were used to define haplotypes (rs3753394, rs529825, rs800292, rs3766404, rs1061170, rs203674, rs3753396, and rs1065489). Of the 36 haplotypes observed at least once, only four haplotypes had an overall frequency greater than 5%; these four haplotypes accounted for 81.9% of the observed haplotypes. We estimated haplotype frequencies and tested association of each haplotype with a frequency of at least 5% in our case–control dataset with age-adjusted score statistics using the haplo.stats software package.[37] We calculated score statistics in the overall AMD dataset (grade 3, 4, and 5 AMD cases compared with grade 1 and 2 controls) and also in a subset of grade 5 neovascular AMD cases compared with grade 1 controls. Empiric probabilities were determined through simulation by using the default simulation options in haplo.stats (*P* precision criterion $P_{\text{threshold}} = 0.25$, so that simulations continued until the sample standard error was less than $P_{\text{threshold}}$ times the probability, minimum number of simulations = 1000, and maximum number of simulations allowed = 20,000). To determine whether the potential protective haplotypes were underrepresented in cases after taking into account the overrepresentation of the risk haplotype in cases, the frequency of the haplotype of interest in cases was compared to the expected frequency of the haplotype given no unequal transmission according to the $\chi^2$ test. This test has been used to show that the frequency of the APOE epsilon 2 allele is reduced in Alzheimer's disease, even after allowing for the increased frequency of the risk APOE epsilon 4 allele in cases of Alzheimer's.[38] A logistic regression model considering P1, P2, and risk haplotypes with the neutral haplotype as the baseline gave similar results. Haplotypes were tested for association with AMD in the family-based dataset by using the HBAT option of the FBAT software package.[39]

In the case–control dataset of all AMD individuals we used the haplo.glm module of haplo.stats to build a regression model to study haplotype–smoking interactions.[40] According to the recommendation of Lake et al.,[40] we modeled only haplotypes with a frequency greater than 5%. Four haplotypes in our dataset met this criterion, and one was selected as the baseline haplotype. Though the selection of the baseline neutral haplotype is somewhat arbitrary, we chose the TGGTTTGT haplotype (Fig. 1) because it was the most frequent haplotype occurring in the same percentage of cases and controls (13.5% in controls and 13.8% in cases), and its association score statistic comparing all AMD individuals (grades 3, 4, and 5) to all controls (grades 1 and 2) indicated neither a protective nor a risk effect (haplotype score = −0.997, $P = 0.311$). Smoking status was obtained from patient ques-

**TABLE 3.** Single-Marker Analysis of the Eight Haplotype-Defining SNPs

| SNP | Associated Allele | Frequency | | $\chi^2$ | *P* |
|---|---|---|---|---|---|
| | | Cases | Controls | | |
| rs3753394 | C | 0.741 | 0.733 | 0.095 | 0.758 |
| rs529825 | G | 0.849 | 0.739 | 24.286 | 8.30E-07 |
| rs800292 | G | 0.851 | 0.747 | 22.87 | 1.73E-06 |
| rs3766404 | T | 0.932 | 0.899 | 4.637 | 0.031 |
| rs1061170 | C | 0.568 | 0.419 | 30.208 | 3.88E-08 |
| rs203674 | G | 0.606 | 0.464 | 25.918 | 3.56E-07 |
| rs3753396 | A | 0.844 | 0.834 | 0.225 | 0.635 |
| rs1065489 | G | 0.841 | 0.829 | 0.289 | 0.591 |

tionnaires and dichotomized by classifying individuals who had smoked >100 cigarettes in their lifetime as "smokers." The full model contained age, smoking, the three most frequent haplotypes (later denoted as protective haplotype P1, protective haplotype P2, and the risk haplotype), a coefficient for the combined effect of the haplotypes with a frequency less than 5%, and four haplotype–smoking interaction terms. The neutral haplotype was used as the referent group in this model. A separate coefficient was used to estimate the effect of the rare haplotypes instead of pooling them with the baseline haplotype, to ensure that any effect of the rare haplotypes would not confound the neutral baseline haplotype. We followed up the P2–smoking interaction with a likelihood ratio test to see whether a model containing the P2–smoking interaction term is significantly different from a model that does not contain this interaction. First, we used a full model containing all haplotypes with a frequency greater than 5%, age, smoking, and the haplotype–smoking interaction terms compared with a reduced model without the four interaction terms. Because we were particularly interested in only the effect of the P2–smoking interaction, we also performed a likelihood ratio test comparing the full model to a reduced model missing only the P2–smoking interaction. Finally, a stratified regression analysis was used to examine the effect of P2 separately in smokers and nonsmokers after adjustment for age.

## RESULTS

### Association of Protective Haplotypes

Two protective haplotypes defined by alleles at eight SNPs in the CFH gene were studied. P1 corresponds to alleles CAATTTAG and is equivalent to H2 from Hageman et al.[19] P2 corresponds to CGGCTTAG and matches H4 from Hageman et al. There are minor differences in the nomenclature between our results and those of Hageman et al. because of differences in which strand of DNA was genotyped. Also, in Figure 3 of Hageman et al. at SNP rs203674 the G and T alleles should be reversed (personal communication, Hancox L, 2005). A risk haplotype CGGTCGAG, corresponding to the H1 haplotype of Hageman et al. 2005 and containing the C allele defining Y402H, was also defined. Figure 1 presents the pair-wise linkage disequilibrium statistics D′ and $r^2$ for these eight SNPs and the haplotype frequencies in the overall dataset, the cases, and the controls. Table 3 describes the results of single-marker tests of association for the eight haplotype-defining SNPs.

Using haplo.stats to calculate age-adjusted score statistics, we found that P1 was significantly associated with AMD in the overall case–control dataset ($P = 0.001$), and when the analysis was restricted to individuals with only the neovascular form of AMD ($P = 0.002$, Table 4). P1 was also significantly overrepresented in the controls, even after taking into account the underrepresentation of the risk haplotype in the controls ($P = 0.016$, Supplementary Table S1, online at http://www.iovs.org/cgi/content/full/48/9/4277/DC1). Logistic regression modeling confirmed this result (P1, $P = 0.014$, in a logistic regression model considering P1, P2, and risk haplotypes with the neutral haplotype as the baseline). The association of P1 was also confirmed in the independent overall family-based dataset ($P = 0.02$, Table 5). Because of the low frequency of this haplotype in the families, it was not possible to get a reliable test of association of P1 in our families exhibiting primarily neovascular AMD.

The age-adjusted score statistic for the P2 haplotype trended toward statistical significance in the overall case–control dataset ($P = 0.077$) and in a restricted analysis of individuals with neovascular AMD ($P = 0.069$, Table 4). In the overall case–control dataset, the P2 haplotype was overrepresented in controls after accounting for the underrepresentation of the risk haplotype ($P < 0.001$, Supplementary Table S1, http://www.iovs.org/cgi/content/full/48/9/4277/DC1). These analyses agreed well with a logistic regression model considering P1, P2, and risk haplotypes with the neutral haplotype as the baseline (P2, $P = 0.059$). The protective effect of the P2 haplotype was present in the overall family-based dataset ($P = 0.018$) and bordered on significance when the analysis was restricted to individuals with neovascular AMD ($P = 0.08$; Table 5).

### Confirmation of the Risk Haplotype

The risk haplotype was very strongly associated with AMD in the overall case–control dataset and in cases with neovascular AMD ($P \leq 1.0^{-5}$ in both analyses; Table 4). The association was confirmed in the overall family-based analysis ($P = 0.04$) and trended toward significance in the family-based dataset

**TABLE 4.** Age-Adjusted Score Statistics for Common Haplotypes in the Independent Case–Control Dataset

| Haplotype | All AMD | | Neovascular AMD | |
|---|---|---|---|---|
| | Score Statistic | *P* | Score Statistic | *P* |
| P1 | −3.254 | 0.001 | −2.975 | 0.002 |
| P2 | −1.780 | 0.077 | −1.794 | 0.069 |
| Risk | 4.451 | 1.0E-05 | 5.144 | <1.0E-05 |

**TABLE 5.** Family-Based Dataset Haplotype Association Results

| Haplotype | Overall Frequency | All AMD *P* | Neovascular AMD *P* |
|---|---|---|---|
| P1 | 0.060 | 0.021 | NA |
| P2 | 0.044 | 0.018 | 0.080 |
| Risk | 0.671 | 0.041 | 0.067 |

Because of the low frequency of P1 in the family-based dataset, none of the families was informative for the HBAT test of association for P1. NA, not available.

when only individuals with neovascular AMD were classified as affected ($P = 0.067$; Table 5).

## Protective Haplotype–Smoking Interaction

Because smoking is a known risk factor for AMD and the frequencies of the haplotypes in cases and controls varied by smoking status (Table 6), we built a regression model incorporating smoking and haplotype–smoking interactions. The full regression model considered age, smoking, haplotype, and haplotype–smoking interactions. The main effects of age, smoking, the risk haplotype, and the pooled rare haplotypes were statistically significant (Table 7). The interactions between P2 and smoking ($P = 0.008$) and the pooled rare haplotypes and smoking ($P = 0.032$) appear to be significant, but a more formal likelihood ratio test is needed for a true assessment of their significance. The likelihood ratio test between the full model with age, smoking, haplotypes with a frequency greater than 5%, and the haplotype–smoking interactions compared to a model without the four haplotype–smoking interaction terms trended toward significance ($\chi^2 = 8.37$, $df = 4$, $P = 0.08$).

Because the frequency of P2 was higher in smoking controls than in smoking cases (10.0% vs. 5.3%), while the frequencies are similar in nonsmokers (4.3% nonsmoking controls vs. 6.4% nonsmoking cases), and because of the low probability for the P2–smoking interaction term in the overall regression model, the P2–smoking interaction is of particular interest. To test specifically for a P2–smoking interaction, we performed another likelihood ratio test, comparing the full model with a reduced model without the P2–smoking interaction. The other haplotype–smoking interaction terms were retained in the reduced model. This test was significant ($\chi^2 = 4.22$, $df = 1$, $P = 0.04$). Furthermore, a stratified regression analysis testing the effect of P2 in smokers versus nonsmokers after adjustment for age and controlling for the effects of the other common haplotypes showed that the protective effect of P2 bordered on significance in smokers ($P = 0.070$), but not in nonsmokers ($P = 0.196$), lending additional confidence to the haplotype–smoking interaction.

**TABLE 7.** Full Regression Model with Haplotype-Smoking Interactions

| Effects | Coefficient | SE | *P* |
|---|---|---|---|
| Intercept | −11.197 | 0.688 | <0.0001 |
| Age | 0.152 | 0.012 | <0.0001 |
| Smoking | 1.083 | 0.441 | 0.014 |
| Protective 1 | −0.157 | 0.322 | 0.626 |
| Protective 2 | 0.760 | 0.489 | 0.121 |
| Risk | 0.558 | 0.233 | 0.017 |
| Pooled rare haplotypes | 0.688 | 0.306 | 0.025 |
| P1-smoking | −0.285 | 0.434 | 0.512 |
| P2-smoking | −1.582 | 0.595 | 0.008 |
| Risk-smoking | −0.101 | 0.311 | 0.745 |
| Pooled rare-smoking | −0.857 | 0.399 | 0.032 |

## DISCUSSION

We have confirmed the presence of protective haplotype P1, not only in a case–control dataset, but also, for the first time, in a family-based dataset. P2 was significantly associated with AMD in the family-based dataset and was marginally significant in the case–control dataset. As expected, the risk haplotype containing the risk allele of Y402H was verified to be strongly associated in both the case–control and family-based datasets. Age-adjusted score statistics for the protective haplotypes in the overall analysis were quite similar to those obtained when only the neovascular AMD cases were studied, suggesting that the protective variant(s) on these haplotypes may be protective for the broad AMD phenotype. Most interesting, a possible protective haplotype–smoking interaction was discovered, though further confirmation of this interaction in independent datasets is necessary to draw firm conclusions.

Although statistical analyses and confirmation in multiple independent datasets of both families and singleton cases and controls has provided strong support for a protective effect of these haplotypes, functional experiments are needed to determine which variant(s) on these haplotypes are exerting the protective influence and the underlying biological mechanism by which they are operating. The CFH protein is composed of 20 short consensus repeats (SCRs) and contains binding sites for complement component C3b, heparin, C-reactive protein (CRP), and sialic acid.[41] CFH functions as a brake on the alternative complement cascade by preventing the activation of C3 to C3a and C3b and by inactivating existing C3b. Sialic acid, heparin, and CRP all facilitate the interaction between CFH and C3b[42–44] and thus lead to a stronger inhibitory signal. Of the eight SNPs defining the protective haplotypes, three are nonsynonymous amino acid substitutions (rs800292 I62V, rs1061170 Y402H, and rs1065489 D936E), one is a synonymous substitution (rs3753396 Q672), one is a putative promoter variant, and three are intronic changes. The valine-to-isoleucine substitution at rs800292 is located within SCR 1 of CFH, which is one of three C3b binding sites within the protein.[41] The putative promoter polymorphism rs3753394 may influence transcription. Thus, both of these variants may have functional consequences for the CFH protein.

However, the protective effect does not seem to be explained by any of the eight polymorphisms alone (Fig. 1). The P1, P2, and risk haplotypes all carry the same allele for the putative promoter variant, and P1 and P2 carry different alleles

**TABLE 6.** Haplotype Frequencies by Smoking Status

| | Smokers | | | Nonsmokers | | |
|---|---|---|---|---|---|---|
| | Overall | Controls | Cases | Overall | Controls | Cases |
| P1 | 0.133 | 0.204 | 0.100 | 0.179 | 0.217 | 0.153 |
| P2 | 0.068 | 0.100 | 0.053 | 0.056 | 0.043 | 0.064 |
| Risk | 0.478 | 0.348 | 0.540 | 0.454 | 0.436 | 0.467 |
| Neutral | 0.131 | 0.132 | 0.130 | 0.135 | 0.143 | 0.129 |

for the I62V, intron 1, and intron 6 polymorphisms, and so none of these variants alone seems plausible as the source of the protective effect. In addition, both protective haplotypes share the alleles TTAG at the last four SNPs of the haplotype, but the A and G alleles at the last two SNPs in P1 and P2 are also found on the risk haplotype, and the T allele at the intron 10 SNP is also found on the neutral haplotype. Therefore, these variants are unlikely to exert a strong protective effect.

As expected, both protective haplotypes carry the nonrisk T allele at the Y402H polymorphism, but several lines of evidence suggest that the absence of the Y402H risk allele is not sufficient to explain the protective effect. First, the nonrisk T allele is also present on the neutral haplotype. Moreover, the frequency of both P1 and P2 was increased in controls, even after allowing for the decreased frequency of the risk haplotype in controls. Finally, in the Japanese the Y402H C allele did not contribute significantly to increased risk for AMD,[45,46] but other risk and protective haplotypes in CFH were present,[32] which suggests that despite the strong single-locus Y402H effect in Caucasians, additional non-Y402H variation in CFH modulates susceptibility to AMD.

This work, as well as the previous report of an interaction between smoking and the *LOC387715* variant rs10490924 on chromosome 10 that modifies AMD risk,[7] demonstrate the importance of considering known lifestyle risk factors for AMD when performing genetic studies. Smokers have decreased plasma levels of CFH,[47] but little is known about the biological function of *LOC387715* and how smoking affects it. Future molecular work aimed at unraveling the complex interplay between smoking, CFH, *LOC387715*, and AMD will surely provide intriguing results.

One concern with haplotype analysis in general is that when using case–control data, the haplotype for an individual is often not directly observed and may have to be inferred using the person's genotype data. The haplo.glm function avoids the need to infer haplotypes directly by using a weighted regression approach that incorporates the posterior probability of an individual's having a particular haplotype into the model.[40] The interpretation of haplo.glm results is more difficult when significant haplotype ambiguity exists. This haplotype ambiguity is reduced as the proportion of individuals in the dataset with high conditional haplotype probabilities increases. In our case, 86% of individuals had posterior probabilities greater than 0.95, suggesting that, at worst, only modest haplotype ambiguity was present.

Multiple testing is a significant issue in genetic studies and cannot be ignored. We limited the problem of multiple testing in this study in several ways. First, we tested only haplotypes with a frequency greater than 5% for association with AMD, rather than test all observed haplotypes. Second, we believe that confirmation of the protective and risk haplotypes in two independent datasets of different types (case–control and family-based) increases confidence that these associations are true-positive results. Finally, only two haplotype–smoking interactions were tested by likelihood ratio methods (an overall test of all haplotypes interacting with smoking and a specific test of the P2–smoking interaction), not all possible haplotype–smoking interactions.

In summary, our data provide support for the presence of two protective haplotypes in the CFH gene. A potential interaction between one of the protective haplotypes and smoking was also identified. Further work is needed to characterize the variant(s) on the haplotypes that are exerting the protective influence and how cigarette smoking may modify this effect.

## *Acknowledgments*

The authors thank the patients, their families, and the control subjects who participated in the study; Monica de la Paz, Molly Klein, Jennifer Caldwell, Ruth Domurath, Katie Haynes, Valerie Mitchell, Maureen Shaw, J. Donald M. Gass, and Jason Galloway, for diligently working to enroll them; the clinics and clinicians who referred individuals for the study: Southern Retina, L.L.C (C. Harris); Vitreo-Retinal Surgeons (M. Duan and C. Devine); Georgia Retina, P.C., and The Retina Group of Washington; and Jason P. Sinwell, MS, for valuable assistance in the use of haplo.stats.

## *References*

1. Berger J, Fine S, Maguire M, eds. *Age-Related Macular Degeneration.* St. Louis: Mosby; 1999.
2. Friedman DS, O'Colmain BJ, Munoz B, et al. Prevalence of age-related macular degeneration in the United States. *Arch Ophthalmol.* 2004;122:564–572.
3. Ambati J, Ambati BK, Yoo SH, Ianchulev S, Adamis AP. Age-related macular degeneration: etiology, pathogenesis, and therapeutic strategies. *Surv Ophthalmol.* 2003;48:257–293.
4. Seddon JM, Chen CA. The epidemiology of age-related macular degeneration. *Int Ophthalmol Clin.* 2004;44:17–39.
5. Thornton J, Edwards R, Mitchell P, Harrison RA, Buchan I, Kelly SP. Smoking and age-related macular degeneration: a review of association. *Eye.* 2005;19:935–944.
6. Tomany SC, Wang JJ, Van LR, et al. Risk factors for incident age-related macular degeneration: pooled findings from 3 continents. *Ophthalmology.* 2004;111:1280–1287.
7. Schmidt S, Hauser MA, Scott WK, et al. Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum Genet.* 2006;78:852–864.
8. Seddon JM, Ajani UA, Mitchell BD. Familial aggregation of age-related maculopathy. *Am J Ophthalmol.* 1997;123:199–206.
9. Klaver CC, Wolfs RC, Assink JJ, van Duijn CM, Hofman A, de Jong PT. Genetic risk of age-related maculopathy: population-based familial aggregation study. *Arch Ophthalmol.* 1998;116:1646–1651.
10. Klein BE, Klein R, Lee KE, Moore EL, Danforth L. Risk of incident age-related eye diseases in people with an affected sibling: The Beaver Dam Eye Study. *Am J Epidemiol.* 2001;154:207–211.
11. Heiba IM, Elston RC, Klein BE, Klein R. Sibling correlations and segregation analysis of age-related maculopathy: the Beaver Dam Eye Study. *Genet Epidemiol.* 1994;11:51–67.
12. Klein ML, Mauldin WM, Stoumbos VD. Heredity and age-related macular degeneration: observations in monozygotic twins. *Arch Ophthalmol.* 1994;112:932–937.
13. Meyers SM. A twin study on age-related macular degeneration. *Trans Am Ophthalmol Soc.* 1994;92:775–843.
14. Meyers SM, Greene T, Gutman FA. A twin study of age-related macular degeneration. *Am J Ophthalmol.* 1995;120:757–766.
15. Gottfredsdottir MS, Sverrisson T, Musch DC, Stefansson E. Age related macular degeneration in monozygotic twins and their spouses in Iceland. *Acta Ophthalmol Scand.* 1999;77:422–425.
16. Hammond CJ, Webster AR, Snieder H, Bird AC, Gilbert CE, Spector TD. Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology.* 2002;109:730–736.
17. 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.
18. Fisher SA, Abecasis GR, Yashar BM, et al. Meta-analysis of genome scans of age-related macular degeneration. *Hum Mol Genet.* 2005; 14:2257–2264.
19. Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci USA.* 2005;102:7227–7232, 2005.
20. Conley YP, Thalamuthu A, Jakobsdottir J, et al. Candidate gene analysis suggests a role for fatty acid biosynthesis and regulation of the complement system in the etiology of age-related maculopathy. *Hum Mol Genet.* 2005;14:1991–2002.

21. Edwards AO, Ritter R III, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. *Science*. 2005;308:421–424.
22. Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science*. 2005;308:419–421.
23. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385–389.
24. Rivera A, Fisher SA, Fritsche LG, et al. 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*. 2005;14:3227–3236.
25. Souied EH, Leveziel N, Richard F, et al. Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. *Mol Vis*. 2005;11:1135–1140.
26. Zareparsi S, Branham KE, Li M, et al. Strong association of the Y402H variant in complement factor H at 1q32 with susceptibility to age-related macular degeneration. *Am J Hum Genet*. 2005;77:149–153.
27. Thakkinstian A, Han P, McEvoy M, et al. Systematic review and meta-analysis of the association between complementary factor H Y402H polymorphisms and age-related macular degeneration. *Hum Mol Genet*. 2006;15:2784–2790.
28. Jakobsdottir J, Conley YP, Weeks DE, Mah TS, Ferrell RE, Gorin MB. Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am J Hum Genet*. 2005;77:389–407.
29. Maller J, George S, Purcell S, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet*. 2006;38:1055–1059.
30. Conley YP, Jakobsdottir J, Mah T, et al. CFH, ELOVL4, PLEKHA1 and LOC387715 genes and susceptibility to age-related maculopathy: AREDS and CHS cohorts and meta-analyses. *Hum Mol Genet*. 2006;15:3206–3218.
31. Gold B, Merriam JC, Zernant J, et al., and the AMA Genetics Clinical Study Group. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet*. 2006;38:458–462.
32. Okamoto H, Umeda S, Obazawa M, et al. Complement factor H polymorphisms in Japanese population with age-related macular degeneration. *Mol Vis*. 2006;12:156–158.
33. Li M, Tmaca-Sonmez P, Othman M, et al. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat Genet*. 2006;38:1049–1054.
34. The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1. *Control Clin Trials.* 1999;20:573–600.
35. Schmidt S, Saunders AM, De La Paz MA, et al. Association of the Apolipoprotein E gene with age-related macular degeneration: possible effect modification by family history, age, and gender. *Mol Vis*. 2000;31:287–293.
36. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–265.
37. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet*. 2002;70:425–434.
38. Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994;7:180–184.
39. Horvath S, Xu X, Lake SL, Silverman SK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol*. 2004;26:61–69.
40. Lake SL, Lyon H, Tantisira K, et al. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered*. 2003;55:56–65.
41. Rodriguez de CS, Esparza-Gordillo J, Goicoechea de JE, Lopez-Trascasa M, Sanchez-Corral P. The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol*. 2004;41:355–367.
42. Fearon DT. Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc Natl Acad Sci USA*. 1978;75:1971–1975.
43. Kazatchkine MD, Fearon DT, Silbert JE, Austen KF. Surface-associated heparin inhibits zymosan-induced activation of the human alternative complement pathway by augmenting the regulatory action of the control proteins on particle-bound C3b. *J Exp Med*. 1979;150:1202–1215.
44. Mold C, Kingzette M, Gewurz H. C-reactive protein inhibits pneumococcal activation of the alternative pathway by increasing the interaction between factor H and C3b. *J Immunol*. 1984;133:882–885.
45. Uka J, Tamura H, Kobayashi T, et al. No association of complement factor H gene polymorphism and age-related macular degeneration in the Japanese population. *Retina*. 2006;26:985–987.
46. Fuse N, Miyazawa A, Mengkegale M, et al. Polymorphisms in complement factor H and hemicentin-1 genes in a Japanese population with dry-type age-related macular degeneration. *Am J Ophthalmol*. 2006;142:1074–1076.
47. Esparza-Gordilla J, Soria JM, Buil A, et al. Genetic and environmental factors influencing the human factor H plasma levels. *Immunogenetics*. 2004;56:77–82.