Associations of *CFHR1–CFHR3* deletion and a *CFH* SNP to age-related macular degeneration are not independent

To the Editor:

Hughes *et al.*[1] suggested that a common deletion of the *CFHR1* and *CFHR3* genes (*CFHR1–3Δ*) is associated with lower risk of age related macular degeneration (AMD) and that the effect is independent from that of the previously described Y402H allele (rs1061170) in the adjacent *CFH* gene[2]. Others have replicated the *CFHR1–3Δ* association[3,4], and this has spurred further research on the function of the *CFHR* gene family[5]. In addition to the Y402H coding variant, we and others have described a second independent *CFH* allele, marked by the rs1410996 intronic SNP[6,7].

Since the *CFH–CFHR1–CFHR3* genomic region containing both of these risk SNPs and *CFHR1–3Δ* has strong linkage disequilibrium (see **Supplementary Fig. 1**) with common haplotypes extending across the entire region[4], we sought to understand the relationship between these AMD associations in a large sample collection. This issue is potentially relevant to atypical hemolytic uremic syndrome (MIM235400), which has also been linked separately to *CFH* alleles and to *CFHR1–3Δ* (ref. 8).

We genotyped *CFHR1–3Δ* and 20 common SNPs within the *CFH* and *CFHR1–CFHR3* region in 711 individuals with visually impairing advanced AMD of AMD and 1041 controls (see **Supplementary Methods**) with the Affymetrix 6.0 chip[9]. This genotyping included the rs10801555 SNP, a close proxy for Y402H ($r^2 = 0.99$ in a subset of 288 genotyped controls), located 1 kb away, and also the rs10737680 SNP, a perfect proxy for the rs1410996 allele ($r^2 = 1$ in Centre d'Etude du Polymorphisme Humain (CEU) HapMap) located 17.5 kb away in the ninth *CFH* intron. *CFHR1–3Δ* frequencies in affected and unaffected individuals were similar to those of Hughes *et al.*[1] and correlated closely with the rs7542235 SNP ($r^2 = 0.98$).

First, we tested each of the 21 markers individually (**Fig. 1a** and **Supplementary Table 1**). We reproduced associations at the *CFH* Y402H allele ($P = 1.5 \times 10^{-39}$ at rs10801555) and the *CFH* rs10737680 allele ($P = 1.8 - 10^{-37}$). We observed more modest evidence of association of *CFHR1–3Δ* ($P = 7.0 \times 10^{-23}$), with 22% frequency in affected individuals compared to 10% in controls.

Second, because Y402H (rs10801555), rs10737680, and *CFHR1–3Δ*, are in linkage disequilibrium (LD) ($D' \geq 0.99$), we used conditional logistic regression to assess whether they independently conferred risk (**Table 1**). A univariate analysis demonstrated significant association to disease for each marker. When we conditioned on Y402H alone, the *CFHR1–3Δ* effect was present (odds ratio 0.58, 95% confidence interval 0.46–0.72, $P = 2 \times 10^{-6}$), as previously reported[1]. However, when we conditioned on rs10737680, the statistical strength of the protective effect of *CFHR1–3Δ* was substantially mitigated (0.72, 0.55–0.95, $P = 0.02$), though not entirely eliminated. At the same time, conditioning on *CFHR1–3Δ* did not mitigate the effect of the Y402H and rs10737680 associations ($P < 1 \times 10^{-13}$). On the basis of these results, we concluded that the previously reported associations at *CFHR1–3Δ* and rs10737680 were not entirely independent.

To better understand the disease association within that locus, we identified common haplotypes of 21 biallelic markers (**Fig. 1b** and **Supplementary Table 2**). A total of seven haplotypes with frequencies >1% accounted for 95.7% of 3,354 chromosomes. The most frequent *H1* haplotype, containing the Y402H risk allele, was present in 59% of chromosomes from affected individuals but only 37% of control chromosomes. For other haplotypes, we calculated the odds ratio of disease association relative to that of *H1*. As previously observed[6], the haplotype risk profiles can be most parsimoniously divided into three groups: high risk (*H1*, odds ratio = 1; reference), intermediate risk (*H2* and *H3*, odds ratio = 0.60, 95% confidence interval (c.i.) 0.50–0.73) and low risk (*H4*, *H5*, *H6* and *H7*, odds ratio = 0.32, 95% c.i. 0.27–0.38). The

Table 1 Conditional logistic regression of *CFH* Y402H, *CFH* rs10737680 and *CFHR1–3Δ*

Logistic regression model	Y402H (rs10801555)			rs10737680			*CFHR1–CFHR3* deletion		
	OR	95% c.i.	P	OR	95% c.i.	P	OR	95% c.i.	P
Single marker model	0.39	0.34–0.46	1.2×10^{-35}	0.38	0.33–0.45	1.6×10^{-32}	0.37	0.30–0.45	6.5×10^{-21}
Conditional on Y402H (rs10801555)	–	–	–	0.58	0.47–0.71	1.1×10^{-7}	0.58	0.46–0.72	2.3×10^{-6}
Conditional on rs10737680	0.55	0.46–0.66	4.5×10^{-10}	–	–	–	0.72	0.55–0.95	0.02
Conditional on *CFHR1–3Δ*	0.47	0.40–0.55	7.7×10^{-21}	0.45	0.37–0.55	1.8×10^{-14}	–	–	–

Measurement of whether each of the three biallelic markers has a significant additive effect on AMD risk. For each marker we present the additive odds ratio (OR), the 95% c.i and the statistical significance of that OR. The rs10737680 SNP is a perfect proxy for the previously associated rs1410996 intronic *CFH* SNP. The first row presents an unconditional univariate analysis for each marker. The next three rows present the effect sizes of each marker after conditioning on each of the markers.

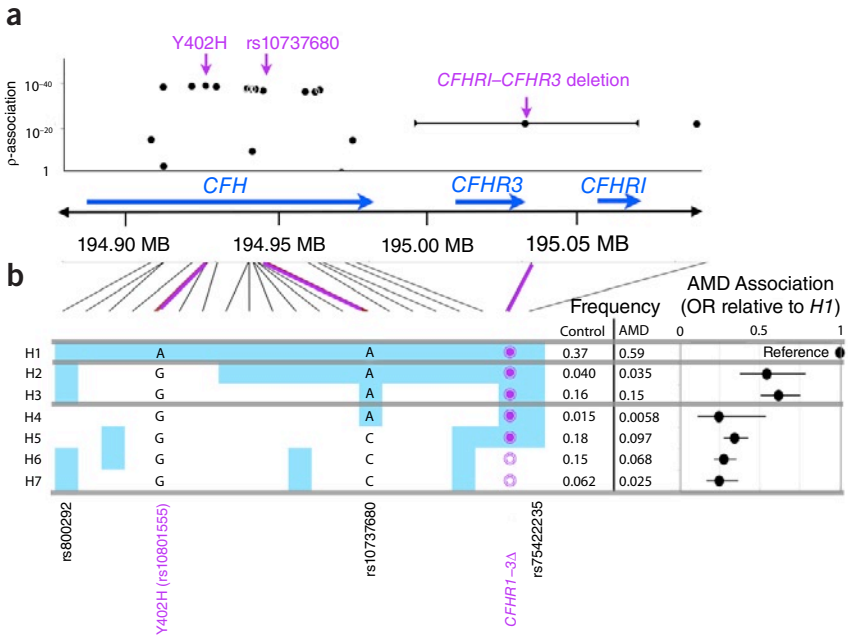

Figure 1 Genetics of the *CFH–CFHR1–CFHR3* region. Statistical results of 20 SNP markers and a *CFHR1–CFHR3* common copy number polymorphism. (**a**) Single marker tests. For each individual marker we plot the statistical strength of association as a function of its genomic position within the region. Violet, previously described SNP associations. (**b**) The seven haplotypes with frequencies >1%. H1 is presented as the reference haplotype. If genotypes for SNPs in other haplotypes are the same as in H1, then they are shaded blue; if genotypes for SNPs differ from H1, they are shaded white. For each haplotype we list the nucleotide for the *CFH* Y402H proxy rs10801555 and for *CFH* rs10737680, and also the deletion status of the *CFHR1–CFHR3* region: empty circle, deleted; filled circle, not deleted. There are two other SNPs of interest: rs7542235, a SNP that tags the *CFHR1–CFHR3* deletion; and rs800292, a *CFH* nonsynonymous (I62V) allele. To the right of each haplotype is the observed frequency in controls and affected individuals. To the far right of each haplotype is the relative ratio of the odds of disease for each haplotype relative to that of the most common haplotype, *H1*.

haplotypes within each group had effect sizes that were indistinguishable from each other ($P = 0.71$ for *H2* and *H3*; $P = 0.30$ for *H4*, *H5*, *H6* and *H7*). The three haplotype groups had distinct effects on AMD risk ($P = 6.8 \times 10^{-43}$), with nonoverlapping confidence intervals; breaking groups to assign independent risk to each of the seven haplotypes did not better define risk ($P = 0.43$).

The haplotype analysis demonstrates the relationship between the *CFH* rs10737680 association and the *CFHR1–3*Δ association: both markers tag a collection of low-risk haplotypes. The rs10737680 SNP is closely linked to the low-risk haplotypes but misses the rare (1.2%) *H4* haplotype, whereas *CFHR1–3*Δ misses both *H4* and *H5*. Neither tags all of the low-risk haplotypes perfectly, suggesting that there could be one or more not-yet-identified variants that better explain disease risk.

One parsimonious explanation is a single protective functional variant present on low-risk haplotypes *H4–H7*, in addition to the Y402H risk allele present on *H1*; such a variant would have very high LD to rs10737680 ($r^2 > 0.9$). Alternatively, a risk variant on intermediate risk haplotypes *H2* and *H3* could also explain the data. We searched for such markers by (i) imputing 171 ungenotyped SNPs with 205 HapMap CEU and Toscani in Italia (TSI) samples as a reference and (ii) imputing 72 ungenotyped *CFH* SNPs with 812 published cases and controls as a reference[7] (**Supplementary Methods**). No genotyped or imputed SNP fulfilled these criteria. Potentially, dense resequencing of this region to ascertain all common variants within this region could identify a functional mutation that fulfills the above criteria.

An alternative but less parsimonious explanation would be the presence of multiple protective functional mutations on the *H4–H7* haplotypes that confer approximately equal effect on risk. For example, *CFHR1–3*Δ or a *CFH* variant in LD on *H6* and *H7* haplotypes and the rs800292 *CFH* coding variant (I62V) on *H4* and *H5* haplotypes might each confer equivalent protection from disease, and this would explain the observed data.

We and others have published examples in which common genomic copy number variation might alter disease risk. For example, the *IRGM* association to Crohn's disease maps to an upstream deletion in the regulatory region, that affects the expression of the gene itself[10]. However, these results suggest the possibility that *CFHR1–3*Δ may not confer any independent risk of AMD, but may simply be associated with protective *CFH* haplotypes.

Soumya Raychaudhuri[1–3], Stephan Ripke[2,3], Mingyao Li[4], Benjamin M Neale[2,3,5], Jesen Fagerness[2,3], Robyn Reynolds[6], Lucia Sobrin[7], Anand Swaroop[8], Gonçalo Abecasis[9], Johanna M Seddon[6,10] & Mark J Daly[2,3]

[1]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA. [2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [3]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [4]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. [5]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK. [6]Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, Massachusetts, USA. [7]Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Boston, Harvard Medical School, Boston, Massachusetts, USA. [8]Neurobiology-Neurodegeneration & Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, Maryland, USA. [9]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [10]Department of Ophthalmology, Tufts Medical Center, Boston, Massachusetts, USA.
e-mail: soumya@broad.mit.edu, mjdaly@chgr.mgh.harvard.edu or jseddon@tuftsmedicalcenter.org

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We appreciate the support of an anonymous donor to the research of J.M.S., without whom this research would not have been possible. We thank the participants and many ophthalmologists throughout the country who contributed to this study, and D. Mirel and the US National Center for Research Resources (NCRR) Broad Institute Center for Genotyping and Analysis for help with design and execution of the genotyping. This research was supported in part by grants K08AR055688-01A1 (S. Raychaudhuri), RO1-EY11309 (J.M.S.), K12-EY16335 (L.S.), U01 MH085520-01 (S. Ripke) and R01-HG004517 (M.L.) from the US National Institutes of Health (NIH); NIH National Eye Institute intramural program; Massachusetts Lions Eye Research Fund, Inc.; a Challenge Grant from Research to Prevent Blindness to the New England Eye Center, Department of Ophthalmology,

Tufts University School of Medicine; a Career Development Award from Research to Prevent Blindness (L.S.); a Harvard Catalyst Faculty Fellowship (L.S.); and the Macular Degeneration Research Fund of the Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Tufts University School of Medicine. We thank the Myocardial Infarction Genetics Consortium (MIGen) study for the use of their genotype data as control data in our study. The MIGen study was funded by grants from the NIH–National Heart, Lung and Blood Institute (R01HL087676) and the NIH-NCRR.

AUTHOR CONTRIBUTIONS

S. Raychaudhuri, J.M.S. and M.J.D. conceived this study, conducted statistical analyses, wrote the initial manuscript and interpreted all results. J.M.S., L.S. and R.R. organized the clinical cohort. S. Raychaudhuri, B.M.N. and J.F. conducted initial processing of the SNP data. S. Ripke, M.L., G.A. and AS imputed missing genotype data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hughes, A.E. *et al. Nat. Genet.* **38**, 1173–1177 (2006).
2. Klein, R.J. *et al. Science* **308**, 385–389 (2005).
3. Hageman, G.S. *et al. Ann. Med.* **38**, 592–604 (2006).
4. Spencer, K.L. *et al. Hum. Mol. Genet.* **17**, 971–977 (2008).
5. Jozsi, M. & Zipfel, P.F. *Trends Immunol.* **29**, 380–387 (2008).
6. Maller, J. *et al. Nat. Genet.* **38**, 1055–1059 (2006).
7. Li, M. *et al. Nat. Genet.* **38**, 1049–1054 (2006).
8. Zipfel, P.F. *et al. PLoS Genet.* **3**, e41 (2007).
9. Neale, B.M. *et al. Proc. Natl. Acad. Sci. USA* **107**, 7395–7400 (2010).
10. McCarroll, S.A. *et al. Nat. Genet.* **40**, 1107–1112 (2008).

Hughes *et al.* reply:

The correspondence by Raychaudhuri *et al.*[1] is in response to our publication in 2006 reporting that a common deletion of *CFHR1* and *CFHR3* is associated with lower risk of age-related macular degeneration (AMD)[2]. In both of these studies, many SNPs were genotyped in individuals with AMD and controls, and risk data based on haplotypes were presented. We are pleased to reply and discuss points relating to the interpretation of genetic association data in this complex genomic region that has high, but incomplete, linkage disequilibrium and contains several related genes in which variation may contribute to AMD risk.

Our group was, to our knowledge, the first to identify a common deletion of *CFHR3* and *CFHR1* associated with the most protective haplotype[2]. This was plausibly functional because of the extremely high homology between *CFHR1* and the final exons of *CFH*, though it is not possible to draw conclusions on function from case-control genetic association studies. Since 2005, when the *CFH* Y402H allele (rs1061170) was shown to be associated strongly with AMD, several groups have reported haplotype risk analysis for the region[2–5]. The data presented by Raychaudhuri *et al.*[1] based on 711 individuals with AMD and 1,041 controls (~45% screened) are reasonably comparable to these studies. All have shown notably consistent risk, neutral and protective haplotypes[6] despite often selecting different SNPs for genotyping and using different haplotype block strategies.

In studies in which many SNPs within the *CFH* gene region have been genotyped, rs2274700, rs1410996, rs10737680, or one of several other markers fully correlated with these SNPs, individually give the most significant *P*-values for association with AMD. The association data from Raychaudhuri *et al.*[1] showing a less significant *P*-value for the deletion ($P = 7.0 \times 10^{-23}$) than for rs10801555 ($P = 1.5 \times 10^{-39}$) or rs10737680 ($P = 1.8 \times 10^{-37}$) is a reflection of allele frequencies rather than effect size, as Table 1 in Raychaudhuri *et al.*[1] shows that all three confer virtually equal effects (with odds ratios ranging from 0.37 to 0.39). The deletion allele frequencies of 22% in controls and 10% in affected individuals found by Raychaudhuri *et al.*[1] compare closely with those found (20% and 8%, respectively) by Hughes *et al.*[2].

Meta-analysis in Hughes *et al.*[2] based on 861 individuals with AMD and 441 screened controls reported an odds ratio of 0.43 (95% confidence interval, 0.33–0.54) for the deletion haplotype. Similar meta-analysis using the same combination of Hageman[3] and Hughes *et al.*[2] data sets for the second protective haplotype tagged by rs800292 (I62V) based on 1,270 individuals with AMD and 612 controls produces an odds ratio of 0.51 (95% confidence interval, 0.42–0.62). Raychaudhuri *et al.*[1] report, for the equivalent haplotypes (H6–H7 and H4–H5), odds ratios of 0.38 and 0.47, respectively, when presented in the same format.

We reanalyzed the Hughes *et al.*[2] data set, first using the model presented by Raychaudhuri *et al.* in Table 1 (ref. 1) with Y402H and counts of haplotypes 4 plus 5 (which are equivalent to the minor allele of rs10737680) and, second, using our preferred model based on three potentially functional elements: Y402H, rs800292 and deletion of *CFHR3–CFHR1* (for which rs6677604 and others can be used as a proxy), and found the latter (less parsimonious but more functionally based) model to be superior. As the two logistic regression models contained different numbers of parameters and were not nested, their fit to the Hughes *et al.*[2] data set was best assessed with the Akaike information criterion (AIC), which adjusts the maximized log likelihood to take account of the number of parameters fitted in the model. The AIC value for the model of Raychaudhuri *et al.*[1] was 576.30, which compares unfavorably with AIC value for the Hughes model of 571.59.

The data from Raychaudhuri *et al.*[1] can support either their view that a functional allele in high correlation with rs10737680 may act on all protective haplotypes or, more likely, that an allele may act on both their H2 and H3 intermediate risk haplotypes. They have not identified any genotyped or imputed SNP with either role.

Along with *CFH*, the previous associations with AMD of variants in *C3*, *CFB* and *CFI* strongly implicate regulation of the alternative complement pathway in the mechanism of susceptibility to AMD. Variation in one or more of these factors may contribute to an individual's risk. The >350-kb region encompassing *CFH* and five expressed *CFH*-related genes is more difficult to analyze than many other regions of complex copy number variation. The effect on AMD risk of the rarer deletion of the *CFHR1–CFHR4* genes[7], and other unconfirmed rearrangements and variations that have not yet been genotyped in either the Raychaudhuri *et al.*[1] or Hughes *et al.*[2] data sets remain to be elucidated.

Although it may be most parsimonious to attempt to model risk with the fewest functional elements, we argue that this is unnecessarily restrictive. Within the *CFH* region, AMD risk in the data set of Raychaudhuri *et al.*[1] can only be predicted with two contributing factors if their rare H4 haplotype is ignored, but full risk information can be extracted from their data by genotyping a minimum of three factors.

Early functional studies support our preferred interpretation of both data sets, with risk based on Y402H, I62V and *CFHR1–CFHR3* deletion status. The Y402H polymorphism affects the specificity and affinity of CFH for glycosaminoglycans on cell surfaces, and possibly also for C-reactive protein (CRP), leading to altered retention of CFH and regulation of complement activity on the retinal surface[8,9]. The I62V polymorphism is a conservative change in CFH; however, there is some evidence that the protective I62 allele may have higher binding affinity for C3b, thus inhibiting proconvertase formation and inactivating fluid and surface-bound