Previously published associations to age related macular degeneration of a common *CFHR1/CFHR3* deletion and a *CFH* intronic SNP are not independent.

Soumya Raychaudhuri[1-3], Stephan Ripke[2,3], Mingyao Li[2,3], 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, 02115, USA.
2. Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
3. Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.
4. Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19103, USA.
5. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AP, UK.
6. Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Department of Ophthalmology, Tufts Medical Center, Boston, MA 02111, USA.
7. Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Boston, Harvard Medical School, Boston, Massachusetts 02114, USA.
8. Neurobiology-Neurodegeneration & Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, Maryland, 20892.
9. Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA.
10. Department of Ophthalmology, Tufts University School of Medicine, Boston, MA 02111.

† Corresponding authors

Supplementary Methods.

Study Sample Descriptions.

The methods employed in this study conformed to the tenets of the Declaration of Helsinki and received approval from the appropriate institutional review boards. Informed consent was signed by all participants. Subjects were recruited through ongoing AMD study protocols as described previously[1-5]. Sample ascertainment and genotyping has also been previously described[1,6,7]. Briefly, cases were defined as patients having either geographic atrophy (advanced dry AMD) or neovascular disease (wet AMD) based on fundus photography and ocular examination (Clinical Age-Related Maculopathy Grading System (CARMS) stages 4 and 5)[8]. Controls from the same study were 60 years of age or older, and were defined as individuals without macular degeneration, and without early or intermediate disease, categorized as stage 1, based on fundus photography and ocular examination.

To expand controls, and add more power to the study, we used controls samples from the MIGEN study[9] as well; these samples were also used in a separate AMD genome-wide study[10]. All samples were unrelated self-described white individuals of European descent.

Genotyping and Quality Control.

We genotyped all of the markers for this study with Affymetrix SNP 6.0 GeneChip at the Broad and National Center for Research Resources (NCRR) Center for Genotyping and Analysis; that experiment is described in detail elsewhere[10]. For that genotyping experiment we collected data

on 906,000 genotyped SNPs and 946,000 CNV probes using the Affymetrix 6.0 GeneChip. We also called common copy number variant genotypes using Birdsuite[11]. We only called CNVs obtaining high-quality Birdsuite call scores (<0.025). We applied strict quality control criteria removing SNPs for low call rates (<99%), failing Hardy-Weinberg test ($p<10^{-3}$), or for case-control differences in call rates ($p<10^{-3}$), or low allele frequencies (1%). After removing individuals with low call rates (<95%), we also used Eigenstrat[12] to identify outlier samples that might contribute to stratification. We observed a λ_{gc} of 1.08 genome-wide, indicating that the samples were generally well matched for population ancestry. All analyses described below were repeated with ten Eigenstrat vector co-variates; results were not substantially affected.

We selected Affymetrix 6.0 SNPs passing quality control within the *CFH/R1/R3* region, defined as 194.88 to 195.09 kb on HG18 (see Figure 1), with minor allele frequencies >5%. In total we identified 20 such SNP markers, including (1) the rs10737680 SNP within a *CFH* intron, which is a perfect proxy for the rs1410996 allele (r^2=1 in CEU HapMap) and (2) the rs10801555 SNP, which is a close proxy for *Y402H* (r^2=0.99 in a subset of 288 genotyped controls). We also extracted genotype calls for CNP147, which represents the common deletion overlapping *CFHR3* and *CFHR1.* All SNPs were re-oriented to the positive strand. We only looked at individuals with non-missing data for all 21 markers, resulting in a final data set of 711 cases and 1041 controls, of which 737 controls taken from the MIGEN study.

All quality control steps and data management was performed using PLINK[13].

Statistical Analysis.

Single marker analysis. For each marker we applied a straight-forward 2x2 single degree of freedom test to assess whether any SNP was significantly associated with disease.

Haplotype analysis. For all markers we constructed haplotypes across the locus with PLINK[13]. We selected all haplotypes with frequencies >1%. For each haplotype we calculated case and control frequencies. We also calculated a 1 degree of freedom association test for each haplotype. To compare relative risk conferred by each haplotype we calculated odds ratios and 95% confidence intervals for each haplotype relative to the most frequent one. To assess whether haplotypes confer differential risk, we use a logistic regression model where each haplotype conferred different additive disease risk, and assessed whether the log-likelihood of that model was significantly improved over a model that assumes that all haplotypes conferred the same risk. We used a similar strategy to test groups of haplotypes, where all haplotypes in the same group are constrained to confer the same degree of additive risk, and compared the log-likelihood of that model to a model that assumes that all haplotypes conferred the same risk. Finally we tested whether the log-likelihood of a model based on grouping haplotypes into high, intermediate, and low risk groups (as described in the main text) was significantly improved by allowing each haplotype to have an independent additive effect. These analyses were performed with the PLINK '--chap' option[13].

Conditional analysis with logistic regression. We used logistic regression to test association for three established *CFH* locus markers: the rs10737680 SNP, the *CFH* Y402H SNP, and the *CFHR1-3* deletion. We assumed an additive model where the log odds of disease was proportional to disease allele counts. We conducted a single marker analysis, without any covariates for each marker. We also conducted a conditional analysis where for each marker, we included an additional marker as a covariate. For each analysis we calculated odds ratios

and 95% confidence intervals, and also *p*-values. These analyses were performed with the PLINK '--logistic' option[13] and also with Octave[14].

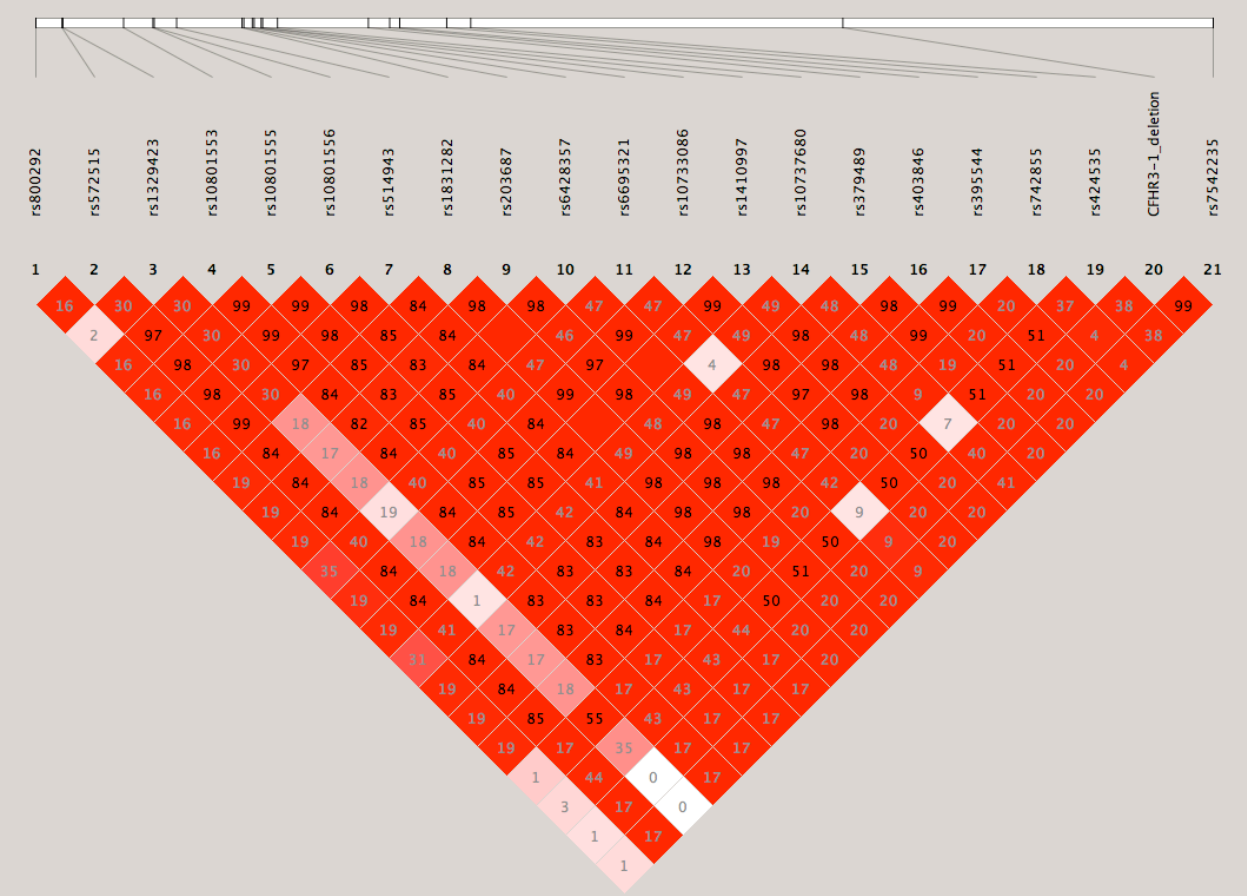
imputation

Using University of Michigan samples as a reference. To impute missing genotypes, we applied MACH[15], a computationally efficient hidden Markov model based algorithm, to impute missing genotypes in the study samples for 72 *CFH* SNPs genotyped in a reference panel of 812 independent samples collected from the University of Michigan, including 544 AMD cases and 268 controls, all of European descent. We first used MACH to phase chromosomes for the Michigan reference samples. Then we applied MACH imputation procedure to study sample genotype data to infer the unknown genotypes probabilistically by searching for similar stretches of flanking haplotype in the reference sample. In this process, we used 10 study SNPs which genotyped in the reference samples. In total, we obtained phased haplotypes at 82 SNPs for the study samples. The estimated mismatch rate for the imputation is 0.00047.

Using HapMap Samples as a reference. To augment genotype information, we applied Beagle[16], to impute genotypes in the study samples using a reference panel of 205 unrelated individuals taken from the phase 3 HapMap CEU and TSI populations (http://hapmap.ncbi.nlm.nih.gov/, version: r2_b36). In this process, we used the 15 study SNPs, along with 2 additional genotyped SNPs either flanking the region of interest (rs17573790 & rs17494275), and 1 SNP excluded from the study for low allele frequencies (rs7513157), that were genotyped in both the study samples and the HapMap samples. Some 5 study SNPs within this study were not genotyped in the HapMap. In total, we obtained probabilistic

genotypes for 171 SNPs across a broad region with high quality scores (r^2>0.4) not genotyped in the study samples.

Supplementary Figure 1.



Supplementary Figure 1. LD across *CFH region.* Data for genetic markers, including 20 SNPs and *CFHR1-3* deletion taken from within the *CFH* region. This plot is based on genetic data from AMD cases and controls, as described in the main text. Degree of red suggests D' between SNPs and values in each box indicates r^2 between SNPs.

Supplementary Table 1.

Marker	Chr 1 Position	Alleles		Disease Allele Frequency (case, control)	OR	*p*-value
		Disease	Protective			
rs800292	194908856	G	A	0.891, 0.787	2.21	1.1E-15
rs572515	194912884	A	G	0.594, 0.369	2.49	4.7E-39
rs1329423	194913010	T	C	0.767, 0.715	1.32	5.0E-04
rs10801553	194922366	A	C	0.596, 0.370	2.51	2.0E-39
*rs10801555 (Y402H proxy)	194926884	A	G	0.597, 0.371	2.51	1.5E-39
rs10801556	194927087	A	G	0.597, 0.371	2.51	1.0E-39
rs514943	194930536	C	T	0.594, 0.369	2.50	3.2E-39
rs1831282	194940616	A	C	0.633, 0.411	2.47	3.5E-38
rs203687	194940893	C	T	0.629, 0.406	2.49	1.3E-38
rs6428357	194942194	A	G	0.633, 0.411	2.47	3.5E-38
rs6695321	194942484	A	G	0.738, 0.635	1.62	1.5E-10
rs10733086	194943558	A	T	0.633, 0.411	2.47	5.0E-38
rs1410997	194943786	C	A	0.633, 0.411	2.47	3.5E-38
*rs10737680	194946078	A	C	0.793, 0.585	2.70	1.8E-37
rs379489	194960074	A	G	0.629, 0.410	2.44	5.7E-37
rs403846	194963360	T	C	0.629, 0.411	2.43	6.9E-37
rs395544	194964895	A	G	0.630, 0.409	2.46	9.7E-38
rs742855	194972143	T	C	0.838, 0.824	1.11	2.8E-01
rs424535	194975846	A	T	0.731, 0.601	1.80	1.9E-15
*CFHR3-1_deletion	194997658 - 195068695	Not Deleted	Deleted	0.904, 0.777	2.72	7.0E-23
rs7542235	195090236	A	G	0.904, 0.777	2.71	9.5E-23

Supplementary Table 1. Single Marker Association Across The *CFH/CFHR3/CFHR1* region. All SNP data presented on the positive strand in HG18. *Previously associated markers (or proxies) examined in this study.

Supplementary Table 2.

Haplotype	Genotype	Frequency (case, control)	OR (relative to Highest risk Haplotype)	*p* (Disease Association)
H1	GATAAACACAAACAATATA◎A	0.592, 0.368	1.00 (REF)	3.5E-38
H2	GGCCGGTACAAACAATATA◎A	0.035, 0.040	0.54 (0.38 - 0.79)	0.46
H3	GGCCGGTCTGGTAAGCGCT◎A	0.154, 0.155	0.62 (0.51 - 0.75)	0.94
H4	AGCCGGTCTGGTAAGCGCT◎A	0.006, 0.015	0.25 (0.11 - 0.54)	1.5E-02
H5	AGTCGGTCTGGTACGCGTA◎A	0.097, 0.176	0.34 (0.28 - 0.43)	1.4E-10
H6	GGTCGGTCTGATACGCGTT◉G	0.068, 0.153	0.28 (0.22 - 0.36)	5.0E-14
H7	GGCCGGTCTGATACGCGTT◉G	0.025, 0.062	0.25 (0.17 - 0.37)	4.0E-07

Supplementary Table 2. Haplotype Association Across The *CFH/CFHR3/CFHR1* region.

For each of the six Haplotypes (column 1) we list the genotype for the 22 markers (see **Supplementary Table 1**) in order. For the *CFHR1-3* deletion a ◎ suggests the presence of a deletion on that haplotype and ◉ and indicates no deletion. We also list the case and control frequencies (column 3); this is identical to the values in Figure 1. We also present the relative odds ratios to the highest risk haplotype, *H1* (column 4). Finally we list the *p*-value of association for each of the individual haplotypes.

References

1. Maller, J. et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* **38**, 1055-9 (2006).
2. Seddon, J.M., Cote, J., Page, W.F., Aggen, S.H. & Neale, M.C. The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch Ophthalmol* **123**, 321-7 (2005).
3. Seddon, J.M., Santangelo, S.L., Book, K., Chong, S. & Cote, J. A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. *Am J Hum Genet* **73**, 780-90 (2003).
4. Seddon, J.M., 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* **121**, 785-92 (2003).
5. Seddon, J.M. et al. Dietary fat and risk for advanced age-related macular degeneration. *Arch Ophthalmol* **119**, 1191-9 (2001).
6. Maller, J.B. et al. Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat Genet* **39**, 1200-1 (2007).
7. Fagerness, J.A. et al. Variation near complement factor I is associated with risk of advanced AMD. *Eur J Hum Genet* **17**, 100-4 (2009).
8. Seddon, J.M., Sharma, S. & Adelman, R.A. Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology* **113**, 260-6 (2006).
9. Kathiresan, S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
10. Neale, B.M. 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* **107**, 7395-400.
11. Korn, J.M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* **40**, 1253-60 (2008).
12. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
13. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
14. Eaton, J. *GNU Octave manual*, (Network Theory Ltd. Bristol, UK, 2002).
15. Li, Y. & Abecasis, G.R. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet* **S79**, 2290 (2006).
16. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* **84**, 210-223 (2009).