Clinical validation of a genetic model to estimate the risk of developing choroidal neovascular age-related macular degeneration

Gregory S. Hageman,[1] Karen Gehrs,[2] Serguei Lejnine,[3] Aruna T. Bansal,[4] Margaret M. DeAngelis,[1] Robyn H. Guymer,[5] Paul N. Baird,[5] Rando Allikmets,[6] Cosmin Deciu,[7] Paul Oeth[7] and Lorah T. Perlee[8*]

[1]Department of Ophthalmology and Visual Sciences, John A. Moran Eye Center, University of Utah, Salt Lake City, UT 84132, USA
[2]The Center for Macula and Retinal Disease, Winter Haven, FL 33880, USA
[3]Bioguidance Consultants, Seattle, WA 98008, USA
[4]Acclarogen Limited, Cambridge CB4 0WS, UK
[5]Centre for Eye Research Australia, Royal Victorian Eye & Ear Hospital, University of Melbourne, East Melbourne, VIC 3002, Australia
[6]Departments of Ophthalmology and Pathology and Cell Biology, Columbia University, New York, NY 10032, USA
[7]Sequenom Center for Molecular Medicine, LLC (SCMM), San Diego, CA 92121, USA
[8]Sequenom, Inc., San Diego, CA 92121, USA
*Correspondence to: Tel: +1 858 202 9050; E-mail: lperlee@sequenom.com

Date received (in revised form): 7th April 2011

Abstract

Predictive tests for estimating the risk of developing late-stage neovascular age-related macular degeneration (AMD) are subject to unique challenges. AMD prevalence increases with age, clinical phenotypes are heterogeneous and control collections are prone to high false-negative rates, as many control subjects are likely to develop disease with advancing age. Risk prediction tests have been presented previously, using up to ten genetic markers and a range of self-reported non-genetic variables such as body mass index (BMI) and smoking history. In order to maximise the accuracy of prediction for mainstream genetic testing, we sought to derive a test comparable in performance to earlier testing models but based purely on genetic markers, which are static through life and not subject to misreporting. We report a multicentre assessment of a larger panel of single nucleotide polymorphisms (SNPs) than previously analysed, to improve further the classification performance of a predictive test to estimate the risk of developing choroidal neovascular (CNV) disease. We developed a predictive model based solely on genetic markers and avoided inclusion of self-reported variables (eg smoking history) or non-static factors (BMI, education status) that might otherwise introduce inaccuracies in calculating individual risk estimates. We describe the performance of a test panel comprising 13 SNPs genotyped across a consolidated collection of four patient cohorts obtained from academic centres deemed appropriate for pooling. We report on predictive effect sizes and their classification performance. By incorporating multiple cohorts of homogeneous ethnic origin, we obtained >80 per cent power to detect differences in genetic variants observed between cases and controls. We focused our study on CNV, a subtype of advanced AMD associated with a severe and potentially treatable form of the disease. Lastly, we followed a two-stage strategy involving both test model development and test model validation to present estimates of classification performance anticipated in the larger clinical setting. The model contained nine SNPs tagging variants in the regulators of complement activation (RCA) locus spanning the complement factor H (CFH), complement factor H-related 4 (CFHR4), complement factor H-related 5 (CFHR5) and coagulation factor XIII B subunit (F13B) genes; the four remaining SNPs targeted polymorphisms

in the complement component 2 (*C2*), complement factor B (*CFB*), complement component 3 (*C3*) and age-related maculopathy susceptibility protein 2 (*ARMS2*) genes. The pooled sample size (1,132 CNV cases, 822 controls) allowed for both model development and model validation to confirm the accuracy of risk prediction. At the validation stage, our test model yielded 82 per cent sensitivity and 63 per cent specificity, comparable with metrics reported with earlier testing models that included environmental risk factors. Our test had an area under the curve of 0.80, reflecting a modest improvement compared with tests reported with fewer SNPs.

Keywords: *age-related macular degeneration (AMD), choroidal neovascularisation (CNV), complement factor H (CFH), genetic testing*

Introduction

Many diseases of ageing characterised by complex inheritance patterns are progressive; the individual may be asymptomatic in the early stages. One of these diseases, age-related macular degeneration (AMD), is the most common cause of visual impairment and the leading cause of blindness in the elderly population in the developed world. The prevalence of AMD increases with advancing age in all populations studied. Thus, in developed nations such as the USA, UK, Canada and Australia, with increasingly aged populations, the condition affects a progressively larger segment of the population and has become a major public health issue. Early- or late-stage AMD is present in 15 per cent of individuals over the age of 60 years.[1] It is estimated that there are currently 9.1 million patients in the USA with AMD, of which 1.7 million suffer with the vision-threatening late-stage complications of choroidal neovascularisation (CNV) or geographic atrophy.[1] Moreover, it is predicted that the number of cases of early AMD will increase to 17.8 million by 2050 and, if untreated, cases of late-stage blinding AMD will increase to 3.8 million.[1] It has been determined that vision loss from AMD decreases quality of life by 60 per cent, similar to the experience of dealing with a stroke that requires intensive nursing care.[2]

The clinical presentation and natural course of AMD are highly variable. The disease may present as early as the fifth decade of life or as late as the ninth decade. The clinical symptoms of AMD range from no visual disturbances in early disease to profound loss of central vision in the advanced late stages of the disease. Some patients never progress beyond early AMD; however, in 10–15 per cent of Caucasian patients with early-stage disease, the condition progresses to an exudative neovascular (or 'wet' form) or geographic atrophic (or 'dry' form) AMD, which threatens vision. The phenotype is characterised by development of subretinal choroidal neovascular complexes, haemorrhage and fibrosis and is typically associated with severe central vision loss.[3,4]

AMD has been one of the success stories of the genome revolution and is probably one of the best characterised of the complex trait diseases in terms of genetic predisposition (for reviews, see Allikmets and Dean[5] and Swaroop *et al.*[6]). Besides age, genetic background is the most significant non-modifiable risk factor for all stages of AMD, while smoking is the most significant modifiable risk factor.[7,8] Initial groundbreaking studies established that loci on chromosomes (Chr) 1 and 10 — in particular the complement factor H (*CFH*) and the age-related maculopathy susceptibility protein 2 (*ARMS2*)/high temperature requirement factor A1 (*HTRA1*) genes, respectively — are significantly associated with AMD risk and protection in populations of various ethnicities.[9–19] Although the specific role(s) of the Chr 10 genes in AMD pathobiology has not yet been elucidated, the role of the alternative complement pathway, where *CFH* functions as a major fluid-phase regulator, is well established (see Anderson *et al.*[20,21] Gehrs *et al.*[22,23] Hageman *et al.*[24,25] and Mullins *et al.*[26] for overviews). Early pathobiological investigations showed dysregulation

of the complement cascade to be a critical early pre-disposing step in the development of AMD. This spurred the discovery of the association of *CFH* variants with AMD risk. Subsequent genetic investigations revealed additional associations between AMD and risk/protective variants in various complement pathway-associated genes, including complement component 2 (*C2*), complement factor B (*CFB*), complement component 3 (*C3*), complement factor H-related 1 and 3 (*CFHR1* and *CFHR3*) and complement factor I (*CFI*).[21,27–38] Using a genome-wide association approach, a handful of additional AMD-associated loci have been reported recently; these appear to be modestly associated with AMD risk and will probably require replication in additional cohorts to establish their role in AMD pathogenesis[39,40] (see also Gehrs *et al.*[23] for a review).

A prerequisite for a new era in genetic testing and diagnosis for AMD is a robust test that accurately captures the impact of consistently replicated AMD risk variants in predicting the risk of developing CNV. Patients with CNV represent an important segment of the AMD population that would benefit from early diagnosis, given the current availability of an effective therapeutic intervention. Jakobsdottir and coworkers[41] recently concluded that the diagnostic value of three variants in the *CFH*, *ARMS2/HTRA1* and *C2* genes was not sufficient to discriminate between individuals with and without AMD because of the relatively low sensitivity and specificity of the combined test panel, in combination with the relatively low prevalence of late-stage disease in the general population. They applied a three single nucleotide polymorphism (SNP) test to their cohort of 640 late-stage AMD cases and 142 controls to demonstrate a clinical sensitivity of 74 per cent and a specificity of 69 per cent, with a reported area under the curve (AUC) — a measure of how well a test or classifier can distinguish between cases and controls — of 0.79. Perfect test discrimination would yield an AUC of 1.0. Jakobsdottir and colleagues also reported that the positive predictive value (PPV) of the same test is affected by different values of disease prevalence reflective of age.

Seddon and colleagues[42] evaluated six AMD risk-associated variants in *CFH, ARMS2/HTRA1, C2, CFB* and *C3* with the goal of developing a predictive risk test for late-stage AMD. After controlling for smoking, body mass index (BMI) and vitamin intake, they demonstrated a strong association between these six risk variants and the prevalence of late-stage AMD, as well as progression to late-stage disease in early AMD patients. The progression test described by Seddon *et al.*,[42] which included genetic, environmental and treatment variables, achieved a performance of 83 per cent sensitivity and 68 per cent specificity, with a reported AUC of 0.82. McKay and co-workers[43] extended this test further, proposing a ten-SNP panel plus smoking history to predict the risk of late-stage AMD. Their inclusion of six CFH SNPs was designed to capture the haplotype structure of the locus, to improve classification performance. Zanke and colleagues[44] have presented risk scores by selecting marker-specific odds ratios from disparate sources and multiplying them together. As the latter approach does not benefit from a joint assessment of the markers (as they perform in combination), it may overestimate an individual's risk of disease.

In this study, we assessed the accuracy of a panel of 13 SNPs without consideration of environmental risk factors such as smoking or BMI, to predict the risk of developing CNV in Caucasian individuals 60 years of age and older. Test model development and validation were designed to evaluate these variants in eight AMD-associated genes (*CFH*, complement factor H-related 4 (*CFHR4*), complement factor H-related 5 (*CFHR5*) and coagulation factor XIII B subunit (*F13B*) located within the regulators of complement activation (RCA) region on Chr 1, *C2* and *CFB* on Chr 6, *C3* on Chr 19 and *ARMS2* on Chr 10. The panel of 13 SNPs was tested in well-established case–control and sibling pair cohorts from five academic centres (University of Iowa, University of Utah, Columbia University, Harvard University and Melbourne University) to validate the accuracy of the predictive test and to estimate an individual's genetic risk for developing late-stage

CNV. Most of the disease-associated genetic variants in *CFH, ARMS2, C2, CFB* and *C3* were selected based on prior replication in multiple studies and performance in resolving the most frequent CFH haplotype combinations. Additional SNPs detecting variants in *CFHR4* (rs1409153), *CFHR5* (rs10922153 and rs1750311) and *F13B* (rs698859 and rs2990510) tagged novel extended haplotypes spanning the *CFH*-to-*F13B* region and were included to maximise the resolution of clinically relevant subtypes suspected to have high association with disease.[45] The additional SNPs were selected to distinguish the novel haplotypes from the more prevalent haplotypes reported previously (H1, H2, H3, H4).[13] The performance metrics obtained during the clinical validation of the 13-SNP panel were used as a benchmark to compare with other published AMD-predictive tests directed at estimating an individual's risk of developing late-stage disease. Since the inclusion of several established non-genetic factors (eg smoking) was highly variable across the published tests, the focus of this investigation was to isolate the contribution conferred by genetic variation alone, in order to determine whether the more comprehensive collection of SNPs could further improve prediction accuracy. The methodology used in the clinical validation of the 13-SNP test panel was subsequently applied to two panels of markers[32,42] that had been assessed previously and contained variants that overlapped with the markers contained within our 13 SNP panel. Both test panels were evaluated in the large collective cohort by using a validation step absent in prior publications. Testing the two panels in a large collection of subjects from different centres assembled from several independent collections was designed to minimise the introduction of selection bias inherent in a single cohort study. Additionally, the use of an independent validation sample was intended to aggressively challenge the 13-SNP panel, to anticipate performance metrics in a broader clinical setting more accurately. Running the three test panels (three SNPs, six SNPs and 13 SNPs) on the same samples allowed for the comparison of performance metrics based exclusively on genetic variants.

Materials and methods

Subjects

Four well-characterised cohorts (Iowa,[13,30] Boston,[38] Columbia,[13,30] and Melbourne[46,47]) and one recently acquired, but as yet unreported, cohort (Utah), together comprised 1,709 patients diagnosed with CNV and 1,473 disease-free controls (for which genotyping data were already available), were assessed (Table 1). All individuals were of white European ancestry, 60 years of age and older and matched for age. All patients had given their consent and were enrolled under Institutional Review Board-approved protocols. The methods used in this study conformed to the tenets of the Declaration of Helsinki (2000) of the World Medical Association. Study subjects were examined and photographed by trained ophthalmologists; fundus photographs were graded according to published standardised classification systems. The worst affected eye of each case was used for classification purposes. All cohorts were case-controlled, with the exception of the Boston sib-pair cohort. Index patients in the Boston cohort aged 60 years or older were included in the analyses and had CNV, (as defined by subretinal haemorrhage, fibrosis or fluorescein angiographic presence of neovascularisation documented at the time of, or prior to, enrolment in the study) in at least one eye. The unaffected siblings had normal maculae at an age older than that at which the index patient was first diagnosed with CNV, as

Table 1. Number of cases (CNV disease) and controls in individual cohorts

Cohort	Control	CNV
Boston	198	338
Columbia	368	522
Iowa	365	284
Melbourne	441	472
Utah	101	93
Total	**1,473**	**1,709**

CNV, choroidal neovascular

previously described.[38] The Utah case–control cohort was recently ascertained at the John A. Moran Eye Center, University of Utah, in Salt Lake City, Utah, USA, in a fashion identical to that of the Iowa cohort.

Markers

Thirteen SNPs, spanning four physically separate genomic loci, were genotyped in all five cohorts (Table 2). One locus spans the *CFH*, *CFHR4*, *CFHR5* and *F13B* genes and comprises nine SNPs; the second consists of two SNPs, one each in *C2* and *CFB*; the third consists of a single SNP in *C3*; and the fourth consists of a single SNP in *ARMS2*. One of the CFH SNPs (rs12144939) included in the panel tags the CFHR3/1 deletion. The 13 SNPs were selected on the basis of the following characteristics: prior published replication, magnitude of estimated effect size and power to resolve clinically relevant haplotypes (CFH).[5–19]

Statistical methods

Previous analyses of each cohort involved standard quality checks and exclusions. Prior to analysis, the consistency of the assignment of the DNA strand used to detect the SNPs was assessed for all available datasets and any inconsistencies resolved. The percentage of missing data and the genotype frequencies were calculated and tabulated for each SNP, both by study (data not shown) and overall (Table 3). No SNPs showed significant deviation from Hardy–Weinberg equilibrium in the control population ($P > 0.05$).

In order to determine the appropriateness of pooling the available cohorts, a chi-squared test of homogeneity of allele frequency was applied to compare frequencies across cohorts. Cohorts or subcohorts found to be a source of a departure from homogeneity of allele frequency (chi square $P < 0.001$) were excluded from the main analysis.

Individuals with CNV were compared with the control group of subjects with no recorded disease. Genotypic multivariate and univariate

Table 2. Single nucleotide polymorphisms employed in first stage

Marker[48]	Chromosome	Base-pair[49] (Build 36.3)	Base-pair[49] (Build 37.1)	Gene
rs1061170	1	194,925,860	196,659,237	*CFH* (exon 9)
rs2274700	1	194,949,570	196,682,947	*CFH* (exon 10)
rs403846	1	194,963,360	196,696,737	*CFH* (intron 14)
rs12144939	1	194,965,568	196,698,945	*CFH* (intron 15)
rs1409153	1	195,146,628	196,880,005	*CFHR4*
rs1750311	1	195,220,848	196,954,225	*CFHR5*
rs10922153	1	195,245,238	196,978,615	*CFHR5*
rs698859	1	195,274,988	197,008,365	*F13B*
rs2990510	1	195,287,281	197,020,658	*F13B*
rs9332739	6	32,011,783	31,903,804	*C2*
rs641153	6	32,022,159	31,914,180	*CFB*
rs10490924	10	124,204,438	124,214,448	*LOC387155 / ARMS2*
rs2230199	19	6,669,387	6,718,387	*C3*

unconditional logistic regression analyses were performed to evaluate the relationships between risk of CNV and the additively coded genotypes (Supplementary Analysis 1). Odds ratios (ORs) and 95 per cent confidence intervals (CIs) were calculated. The full 13-SNP panel was evaluated both with and without demographic factors of age and sex. Backward elimination was performed on the training set using a threshold of $P < 0.05$.

Table 3. Homogeneity of variance

Cohort	Counts (row frequency) rs10490924 Code = CNTL			Total
	GG	GT	TT	
Boston	101	71	26	198
	51.01%	35.86%	13.13%	100.00%
Columbia	218	136	14	368
	59.24%	36.96%	3.80%	100.00%
Iowa	230	117	13	360
	63.89%	32.50%	3.61%	100.00%
Melbourne	277	145	16	438
	63.24%	33.11%	3.65%	100.00%
Utah	62	39	0	101
	61.39%	38.61%	0.00%	100.00%
Total	888	508	69	1,465

Cohort	Counts (row frequency) rs403846 Code = CNTL			Total
	AA	AG	GG	
Boston	41	102	55	198
	20.71%	51.52%	27.78%	100.00%
Columbia	32	164	165	361
	8.86%	45.43%	45.71%	100.00%
Iowa	68	179	118	365
	18.63%	49.04%	32.33%	100.00%
Melbourne	71	229	137	437
	16.25%	52.40%	31.35%	100.00%
Utah	13	61	27	101
	12.87%	60.40%	26.73%	100.00%
Total	225	735	502	1,462

Continued

Table 3. Continued

Cohort	Counts (row frequency) rs1409153 Code = CNTL			Total
	AA	AG	GG	
Boston	67	97	34	198
	33.84%	48.99%	17.17%	100.00%
Columbia	177	161	29	367
	48.23%	43.87%	7.90%	100.00%
Iowa	128	177	60	365
	35.07%	48.49%	16.44%	100.00%
Melbourne	150	226	63	439
	34.17%	51.48%	14.35%	100.00%
Utah	31	60	10	101
	30.69%	59.41%	9.90%	100.00%
Total	553	721	196	1,470

Cohort	Counts (row frequency) rs10922153 Code = CNTL			Total
	GG	GT	TT	
Boston	53	102	43	198
	26.77%	51.52%	21.72%	100.00%
Columbia	55	181	122	358
	15.36%	50.56%	34.08%	100.00%
Iowa	99	172	94	365
	27.12%	47.12%	25.75%	100.00%
Melbourne	94	234	113	441
	21.32%	53.06%	25.62%	100.00%
Utah	20	59	21	100
	20.00%	59.00%	21.00%	100.00%
Total	321	748	393	1,462

Continued

Table 3. Continued

| Cohort | Counts (row frequency) | | | Total |
| | rs403846 Code = CNV | | | |
	AA	AG	GG	
Boston	141	149	48	338
	41.72%	44.08%	14.20%	100.00%
Columbia	148	255	116	519
	28.52%	49.13%	22.35%	100.00%
Iowa	110	137	37	284
	38.73%	48.24%	13.03%	100.00%
Melbourne	179	218	74	471
	38.00%	46.28%	15.71%	100.00%
Utah	33	46	14	93
	35.48%	49.46%	15.05%	100.00%
Total	611	805	289	1,705

| Cohort | Counts (row frequency) | | | Total |
| | rs698859 Code = CNV | | | |
	AA	AG	GG	
Boston	85	147	105	337
	25.22%	43.62%	31.16%	100.00%
Columbia	78	238	205	521
	14.97%	45.68%	39.35%	100.00%
Iowa	69	136	79	284
	24.30%	47.89%	27.82%	100.00%
Melbourne	76	233	163	472
	16.10%	49.36%	34.53%	100.00%
Utah	19	49	25	93
	20.43%	52.69%	26.88%	100.00%
Total	327	803	577	1,707

Two published test models containing, respectively, three and six SNPs, and a nine-SNP model generated from backward elimination, were compared with the 13-SNP panel in terms of AUC in training and independent validation. In the event that an SNP was not present in the 13-SNP panel, a SNP with demonstrated linkage disequilibrium was used as a surrogate.

Training of classifiers was performed using 500 cases and 500 controls balanced by age and sex and randomly selected from the whole cohort. The remaining 322 controls and 632 cases were used for validation. In both analyses, ten-fold cross-validation was applied.[50] The predicted probability of affliction for each subject was calculated by applying the inverse-logit function; sensitivity, specificity and AUC were derived to assess classification performance.

A risk score for CNV was calculated as follows: $Sj = \text{intercept} + \sum_{i=1}^{13} \beta i * Xi$ where Sj is the risk score for subject j and βi is the adjusted log-odds ratio for Xi, the additively coded genotype at marker i. The probability of risk for subject j was calculated as $pj = \exp(Sj)/[1 + \exp(Sj)]$.

The optimal classification threshold was determined on the basis of accuracy, defined as the proportion of correct predictions observed in cases and controls. Different levels of prevalence, reflecting age-specific differences, were considered. The accuracy in the validation set was determined, and positive and negative predictive values were calculated. Calibration was assessed graphically as histograms showing disease incidence at different levels of predicted risk for controls and cases.

The area under the receiver operating characteristic (ROC) curve and CIs were estimated using SAS Macro %ROC.[51] In addition, c-statistics and CIs were calculated for the training, tenfold cross-validation and validation datasets.[52,53]

All analyses were conducted using SAS 9.1.[52]

Results

The average ages ($\pm$ standard deviation [SD]) of cases and controls among all cohorts were 76.4 (± 7.3) and 76.5 (± 7.1) years, respectively, and the differences were not significant ($p = 0.86$). Age matching was applied during cohort ascertainment. The chi-square test was used to assess homogeneity of allele frequency across cohorts. Frequencies of markers rs10490924, rs403846, rs1409153,

Table 4. Univariate association between demographic, genetic factors and risk of choroidal neovascular (CNV) disease

		Control (822)	CNV (1132)	Odds (95% CI)	*P*-value (Type 3)	c-statistic
Age (± SD)		76.4 (7.3)	76.5 (7.1)	1.001 (0.989–1.013)	0.87	0.50
Sex	F	451 (55%)	696 (61%)	1.313 (1.094–1.576)	0.0034	0.53
	M	371 (45%)	436 (39%)			
rs10490924	GG	520 (63.3%)	340 (30%)	0.061 (0.04–0.093)	< 0.0001	0.70
	GT	269 (32.7%)	505 (44.6%)	0.175 (0.114–0.268)		
	TT	26 (3.2%)	279 (24.6%)			
	(blank)	7 (0.9%)	8 (0.7%)			
rs1061170	CC	114 (13.9%)	394 (34.8%)	5.184 (3.934–6.831)	< 0.0001	0.65
	CT	408 (49.6%)	535 (47.3%)	1.967 (1.575–2.456)		
	TT	294 (35.8%)	196 (17.3%)			
	(blank)	6 (0.7%)	7 (0.6%)			
rs10922153	GG	189 (23%)	498 (44%)	4.819 (3.64–6.382)	< 0.0001	0.64
	GT	418 (50.9%)	515 (45.5%)	2.254 (1.738–2.922)		
	TT	214 (26%)	117 (10.3%)			
	(blank)	1 (0.1%)	2 (0.2%)			
rs12144939	GG	504 (61.3%)	930 (82.2%)	7.996 (3.842–16.639)	< 0.0001	0.61
	GT	275 (33.5%)	192 (17%)	3.025 (1.432–6.391)		
	TT	39 (4.7%)	9 (0.8%)			
	(blank)	4 (0.5%)	1 (0.1%)			
rs1409153	AA	282 (34.3%)	192 (17%)	0.203 (0.154–0.267)	< 0.0001	0.64
	AG	420 (51.1%)	539 (47.6%)	0.382 (0.3–0.487)		
	GG	118 (14.4%)	396 (35%)			
	(blank)	2 (0.2%)	5 (0.4%)			
rs1750311	AA	95 (11.6%)	53 (4.7%)	0.289 (0.202–0.415)	< 0.0001	0.59
	AC	373 (45.4%)	411 (36.3%)	0.572 (0.472–0.692)		
	CC	346 (42.1%)	667 (58.9%)			
	(blank)	8 (1%)	1 (0.1%)			
rs2230199	CC	521 (63.4%)	621 (54.9%)	0.447 (0.289–0.691)	< 0.0001	0.55
	CG	267 (32.5%)	428 (37.8%)	0.601 (0.385–0.94)		
	GG	30 (3.6%)	80 (7.1%)			
	(blank)	4 (0.5%)	3 (0.3%)			

Continued

Table 4. Continued

		Control (822)	CNV (1132)	Odds (95% CI)	P-value (Type 3)	c-statistic
rs2274700	AA	144 (17.5%)	48 (4.2%)	0.128 (0.09–0.183)	< 0.0001	0.66
	AG	403 (49%)	378 (33.4%)	0.361 (0.296–0.441)		
	GG	268 (32.6%)	696 (61.5%)			
	(blank)	7 (0.9%)	10 (0.9%)			
rs2990510	GG	78 (9.5%)	183 (16.2%)	2.082 (1.541–2.813)	< 0.0001	0.55
	GT	389 (47.3%)	544 (48.1%)	1.241 (1.023–1.506)		
	TT	355 (43.2%)	400 (35.3%)			
	(blank)	(0%)	5 (0.4%)			
rs403846	AA	137 (16.7%)	445 (39.3%)	5.059 (3.848–6.652)	< 0.0001	0.65
	AG	424 (51.6%)	521 (46%)	1.914 (1.515–2.418)		
	GG	257 (31.3%)	165 (14.6%)			
	(blank)	4 (0.5%)	1 (0.1%)			
rs641153	CC	644 (78.3%)	984 (86.9%)	2.674 (1.115–6.41)	< 0.0001	0.55
	CT	159 (19.3%)	129 (11.4%)	1.42 (0.578–3.489)		
	TT	14 (1.7%)	8 (0.7%)			
	(blank)	5 (0.6%)	11 (1%)			
rs698859	AA	120 (14.6%)	235 (20.8%)	1.644 (1.257–2.15)	0.0012	0.54
	AG	403 (49%)	541 (47.8%)	1.127 (0.922–1.378)		
	GG	298 (36.3%)	355 (31.4%)			
	(blank)	1 (0.1%)	1 (0.1%)			
rs9332739	CC	2 (0.2%)	1 (0.1%)	0.348 (0.032–3.85)	0.0022	0.52
	CG	72 (8.8%)	55 (4.9%)	0.532 (0.37–0.766)		
	GG	745 (90.6%)	1069 (94.4%)			
	(blank)	3 (0.4%)	7 (0.6%)			

CI, confidence interval

rs698859, rs403846 and rs10922153 were significantly different ($P < 0.001$) across cohorts. The frequencies of four markers — rs10490924, (ARMS2) rs403846, (CFH) rs1409153 (CFHR4) and rs10922153 (CFHR5) — in the control population and two markers — rs698859 (F13B) and rs403846 (CFH) — in the CNV population were unbalanced (Table 3). Removal of the Columbia University cohort eliminated four of the five deviations, leaving only one SNP (rs10490924) outstanding in the Boston control population. The Boston controls and Columbia cases and controls were excluded from the main analyses based on these observations. The remaining study population contained 1,132 CNV cases and 822 controls. For the purposes of the current analysis, investigations into

Table 5. Calculation of choroidal neovascular disease risk score: $S = \text{intercept} + \sum_{i=1}^{13} \beta i * Xi$, where β and X are as follows

Parameter	Regression coefficient	p-value	X	Point estimate	95% Wald confidence limits		Pr[a] > Chisq
Intercept	0.7851	0.1885	1	—	—	—	—
rs10490924	1.4537	<0.0001	GG = 0,GT = 1,TT = 2	4.279	3.346	5.472	<0.0001
rs1061170	−0.7687	0.0105	CT = 1,CC = 0,TT = 2	0.464	0.257	0.835	0.0105
rs10922153	−0.6018	0.1129	GT = 1,GG = 0,TT = 2	0.548	0.26	1.153	0.1129
rs12144939	−0.1974	0.4375	GG = 0,GT = 1,TT = 2	0.821	0.499	1.351	0.4375
rs1409153	−0.1595	0.5665	AG = 1,GG = 0,AA = 2	0.853	0.494	1.471	0.5665
rs1750311	−0.1316	0.6834	CC = 0,AC = 1,AA = 2	0.877	0.466	1.65	0.6834
rs2230199	0.428	0.0009	CC = 0,CG = 1,GG = 2	1.534	1.192	1.975	0.0009
rs2274700	−0.7954	0.0002	GG = 0,AG = 1,AA = 2	0.451	0.296	0.689	0.0002
rs2990510	−0.4596	0.1358	GT = 1,TT = 0,GG = 2	0.632	0.345	1.155	0.1358
rs403846	0.8131	0.0404	AG = 1,AA = 0,GG = 2	2.255	1.036	4.906	0.0404
rs641153	−0.8243	<0.0001	CC = 0,CT = 1,TT = 2	0.439	0.295	0.651	<0.0001
rs698859	−0.015	0.9559	AG = 1,GG = 0,AA = 2	0.985	0.58	1.673	0.9559
rs9332739	−0.9544	0.0027	GG = 0,CG = 1,CC = 2	0.385	0.206	0.719	0.0027

[a] The probability of risk = exp(risk score)/[1 + exp(risk score)]

the differences were not pursued but could be evaluated in the future by performing structure analysis to identify potential causes for the observed differences.

Table 4 shows unadjusted association test results between the demographic and genetic factors and the risk of CNV. All factors except age were associated with risk of CNV. The c-statistic column shows the ability of a genetic factor to predict CNV risk. SNPs rs10490924, rs1061170, rs403846 and rs2274700 had c-statistics ≥ 0.65.

Table 5 displays multivariate adjusted ORs that were significantly associated with the risk of CNV, using the additive genotype model applied to the 13-SNP panel. The ARMS2 variant rs10490924 was positively associated with risk of CNV (OR 4.279, 95 per cent CI 3.346−5.472, $p < 0.0001$).

The performance of the 13-SNP panel to predict CNV relative to the control population was evaluated using tenfold cross-validation and an independent dataset. Independent datasets were scored using model parameters displayed in Table 5. Table 6 shows the AUC evaluated for training (AUC 0.82 [0.79−0.85]), tenfold cross-validation (AUC 0.81 [0.79−0.84]) and validation (AUC 0.79 [0.77−0.83]). The c-statistics results were identical to AUC. These data show that the difference in

Table 6. Area under the curve for training, tenfold cross-validation and independent validation on 13-SNP model

Stage	Control/CNV	ROC area	Standard error	Confidence limits	
Training	467/482	0.82	0.01	0.79	0.85
Tenfold cross-validation	467/482	0.81	0.01	0.79	0.84
Validation	322/632	0.80	0.02	0.77	0.83

SNP, single nucleotide polymorphism; CNV, choroidal neovascular; ROC, receiver operating characteristic

Table 7. Comparison of 13-SNP model with and without demographic factors. There is no significant difference between the two models

| Step	Model	ROC area	Standard error	Confidence limits
Training	Age + Sex + 13 SNP	0.82	0.01	0.79–0.85
Training	13 SNP	0.82	0.01	0.79–0.85
Validation	Age + Sex + 13 SNP	0.80	0.02	0.77–0.83
Validation	13 SNP	0.80	0.02	0.77–0.83

ROC, receiver operating characteristic; SNP, single nucleotide polymorphism
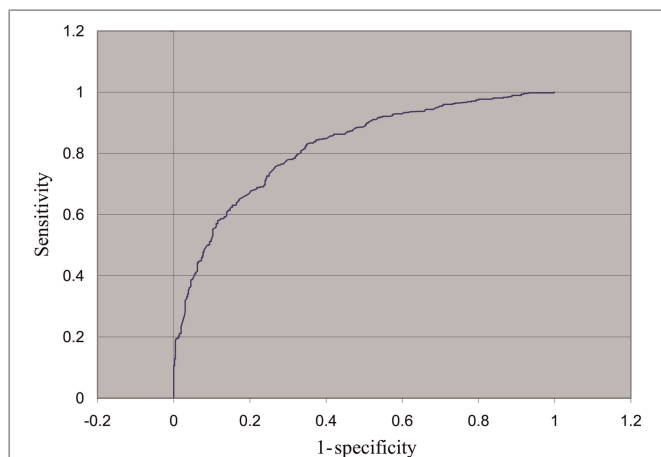
Figure 1. ROC curve for validation.
ROC, receiver operating characteristic

performance of the training and validation sets was not significant ($P < 0.05$). There were no significant differences between the AUC curves for the training and validation datasets with demographic factors (age, sex) added into the test model (Table 7), presumably due to the balanced study design.

The sensitivity and specificity of predictions were calculated in an independent dataset using the test panels in Table 5. The ROC curve is shown in Figure 1. The probability of the risk of CNV was plotted as histograms for controls and cases in the independent dataset in Figure 2. It shows good separation between the two groups, with cases having
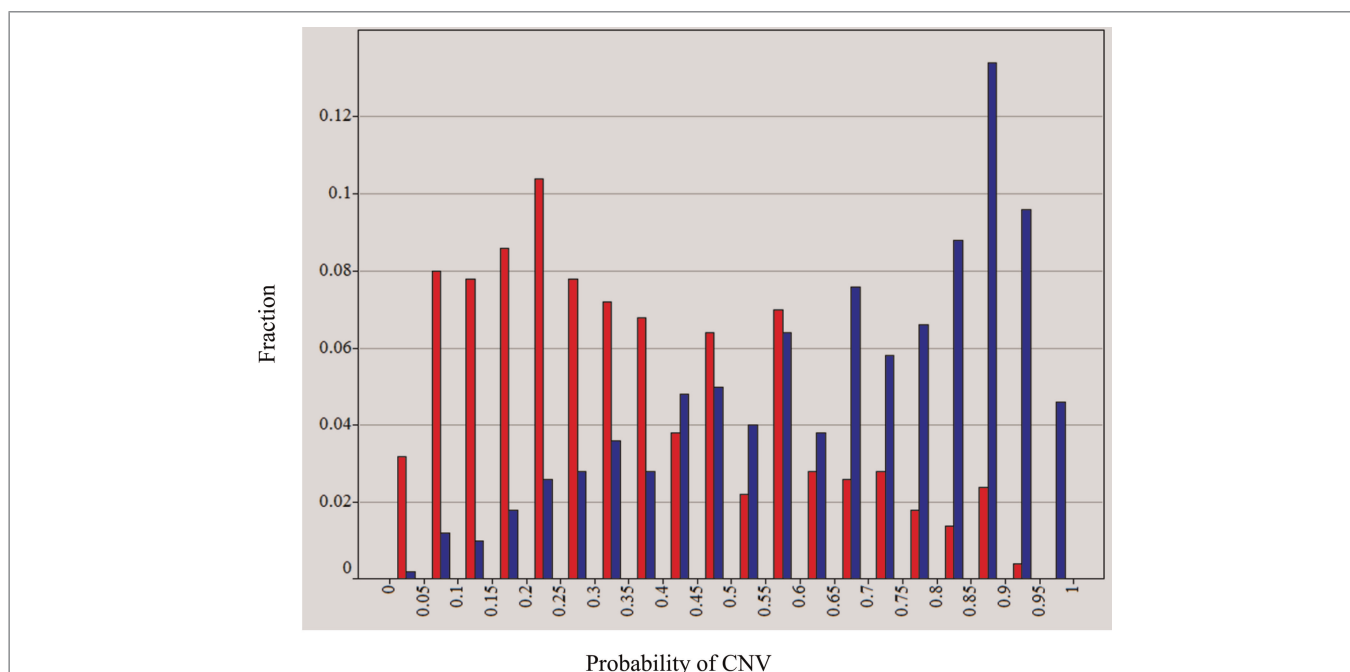
Figure 2. Probability of choroidal neovascular (CNV) disease, calculated for validation dataset using model described in Table 2. Red bars represent controls and blue bars represent patients with CNV disease.

Table 8. Classification table

Prob. level	Sensitivity	Specificity	PPV % (5.5%)	NPV % (5.5%)	PPV % (10%)	NPV % (10%)	PPV % (15%)	NPV % (15%)
0.00	100.0	0.0	5.5	—	10.0	—	15.0	—
0.02	99.8	0.2	5.5	94.5	10.0	90.0	15.0	85.0
0.04	99.8	2.1	5.6	99.4	10.2	99.0	15.2	98.3
0.06	99.8	4.3	5.7	99.7	10.4	99.5	15.5	99.2
0.08	98.8	8.6	5.9	99.2	10.7	98.5	16.0	97.6
0.10	98.1	12.0	6.1	99.1	11.0	98.3	16.4	97.3
0.12	97.7	15.0	6.3	99.1	11.3	98.3	16.9	97.4
0.14	97.3	18.2	6.5	99.1	11.7	98.4	17.3	97.4
0.16	96.7	20.8	6.6	99.1	11.9	98.3	17.7	97.3
0.18	95.9	23.8	6.8	99.0	12.3	98.1	18.2	97.0
0.20	95.0	29.1	7.2	99.0	13.0	98.1	19.1	97.1
0.22	93.6	33.0	7.5	98.9	13.4	97.9	19.8	96.7
0.24	92.9	38.1	8.0	98.9	14.3	98.0	20.9	96.8
0.26	91.7	43.3	8.6	98.9	15.2	97.9	22.2	96.7
0.28	90.5	45.2	8.8	98.8	15.5	97.7	22.6	96.4
0.30	88.8	48.8	9.2	98.7	16.2	97.5	23.4	96.1
0.32	86.9	50.7	9.3	98.5	16.4	97.2	23.7	95.6
0.34	86.1	53.7	9.8	98.5	17.1	97.2	24.7	95.6
0.36	85.5	56.7	10.3	98.5	18.0	97.2	25.8	95.7
0.38	83.4	60.4	10.9	98.4	19.0	97.0	27.1	95.4
0.40	81.7	63.2	11.4	98.3	19.8	96.9	28.1	95.1
0.42	80.5	65.3	11.9	98.3	20.5	96.8	29.0	95.0
0.44	78.4	66.6	12.0	98.1	20.7	96.5	29.3	94.6
0.46	77.8	68.1	12.4	98.1	21.3	96.5	30.1	94.6
0.48	73.7	71.7	13.2	97.9	22.4	96.1	31.5	93.9
0.50	72.4	74.7	14.3	97.9	24.1	96.1	33.6	93.9
0.52	70.3	75.4	14.3	97.8	24.1	95.8	33.5	93.5
0.54	68.9	76.0	14.3	97.7	24.2	95.7	33.6	93.3
0.56	68.5	76.9	14.7	97.7	24.8	95.6	34.4	93.3
0.58	63.9	79.9	15.6	97.4	26.1	95.2	35.9	92.6
0.60	61.4	84.6	18.8	97.4	30.7	95.2	41.3	92.5

Continued

Table 8. Continued

Prob. level	Sensitivity	Specificity	PPV % (5.5%)	NPV % (5.5%)	PPV % (10%)	NPV % (10%)	PPV % (15%)	NPV % (15%)
0.62	60.4	85.4	19.4	97.4	31.5	95.1	42.2	92.4
0.64	58.3	86.1	19.6	97.3	31.8	94.9	42.5	92.1
0.66	56.6	87.6	21.0	97.2	33.7	94.8	44.6	92.0
0.68	51.5	89.1	21.6	96.9	34.4	94.3	45.5	91.2
0.70	50.0	90.4	23.3	96.9	36.7	94.2	47.9	91.1
0.72	47.7	91.4	24.4	96.8	38.1	94.0	49.5	90.8
0.74	44.6	92.3	25.2	96.6	39.2	93.7	50.5	90.4
0.76	43.8	92.9	26.4	96.6	40.7	93.7	52.1	90.4
0.78	41.3	93.8	27.9	96.5	42.5	93.5	54.0	90.1
0.80	37.1	95.1	30.6	96.3	45.7	93.2	57.2	89.5
0.82	33.6	95.7	31.3	96.1	46.5	92.8	58.0	89.1
0.84	30.1	96.4	32.7	96.0	48.2	92.5	59.6	88.7
0.86	22.4	97.9	38.3	95.6	54.2	91.9	65.3	87.7
0.88	20.3	98.1	38.3	95.5	54.3	91.7	65.3	87.5
0.90	14.7	99.6	68.1	95.3	80.3	91.3	86.6	86.9
0.92	10.4	99.8	75.2	95.0	85.2	90.9	90.2	86.3
0.94	7.9	100.0	100.0	94.9	100.0	90.7	100.0	86.0
0.96	3.9	100.0	100.0	94.7	100.0	90.4	100.0	85.5
0.98	0.6	100.0	100.0	94.5	100.0	90.1	100.0	85.1
1.00	0.0	100.0	–	94.5	–	90.0	–	85.0

Prob., probability; PPV, positive predictive value; NPV, negative predictive value

a substantially higher probability of CNV, although some overlap is present.

Accuracy, specificity, sensitivity, PPV and negative predicted values (NPV) are shown in Table 8 as a function of probability cut-off and three prevalence values. A cut-off of 0.4 corresponds to the highest accuracy (0.73), with a sensitivity of 0.82 and a speci–ficity of 0.63. The PPV for 5.5 per cent, 10 per cent and 15 per cent prevalence values were 0.11, 0.20 and 0.28, respectively. The NPVs were all above 0.95.

We compared several published predictive models with our current 13-SNP panel (Table 9). The differences in test performance were evaluated at training and validation stages. The performance of

the 13-SNP panel was slightly better than that of the next best test.[41,42] Results from the nine-SNP panel generated from the backwards elimination procedure realised gains in genotyping efficiency, with four fewer variants in the panel, while demonstrating only slightly lower performance in terms of AUC.

Discussion

Although the incorporation of non-static and self–reported variables is important in elucidating the modifiable risk factors that contribute to disease, their inclusion can degrade test performance in mainstream genetic testing. Ideally, a robust test

Table 9. Comparison of models containing different numbers of single nucleotide polymorphisms (SNPs)

Model	Reported AUC	Current study training AUC	Significance to 13 SNP SCMM training	Current study validation AUC	Significance to 13 SNP SCMM validation
Three-SNP (Jakobsdottir[41])	0.79	0.77	<0.0001	0.77	<0.001
Six-SNP (Seddon[42])	0.82[a]	0.81	<0.01	0.79	<0.05
Nine-SNP (SCMM)	NA	0.81	<0.01	0.79	ns[b]
13-SNP (SCMM)	NA	0.82	—	0.80	—

AUC, area under the curve; SCMM, Sequenom Center for Molecular Medicine.
[a]AUC value based on model with six SNPs and multiple environmental risk variables (eg baseline grade, education status, BMI, smoking history).
[b]ns: not significant ($p > 0.05$).

panel, subject to rigorous validation, which captures the maximal genetic component should improve classification performance and accuracy of reporting. In line with these criteria, which are much stricter than in a discovery cohort, the Boston cohort controls and the Columbia cohort cases and controls were not considered for the calculation of the model. Possible explanations for the allele frequency deviations in these cohorts include admixture, cryptic population stratification, subtle differences in grading criteria, cohort age range, concomitant illnesses or medications, and should be explored further.

In order to compare performance across tests, a ROC curve was generated for each prediction panel to evaluate the AUC. By evaluating each test across the large collective cohort using the same validation procedure, we compared the power of the genetic variants to evaluate classification performance. The performance of the three-SNP panel described by Jakobsdottir and colleagues[41] revealed an AUC value of 0.77, compared with a value of 0.79 observed in the original study of 642 late-stage AMD cases and 142 controls. The differences in AUC values obtained between the original and the current study are likely to reflect the impact of testing across a large collection of independently collected cohorts compared with a single study that is potentially more sensitive to subject selection bias. The performance of the six-SNP test panel reported by

Seddon and colleagues[42] as part of a joint gene–environment model exhibited a drop in AUC from 0.81 to 0.79 from training to validation in our data (significant at $P < 0.05$), similar to most of the tests evaluated. This decrease in AUC reveals the value of the inclusion of an independent validation set to challenge test performance and estimate metrics achievable in the broader clinical setting more accurately. We have emphasised the importance of both study design features to report performance more accurately and to anticipate utility in the more diverse clinical testing market more closely. Finally, modest gains in our 13-SNP panel were demonstrated with the highest AUC value obtained among all models evaluated (0.80). The additional variants that contributed to the performance of the predictive test located in *CFHR5* and F13B highlight the complexity of the genetic structure of the RCA region and influence AMD disease biology.

In summary, the 13-SNP panel had a clinical sensitivity of 82 per cent and a specificity of 63 per cent, achieving clinical performance metrics comparable with models with fewer SNPs that include self-reported and/or non-static risk factors. The PPV of the panel was evaluated at different levels of prevalence, reflecting ranges covering estimates of late-stage disease in individuals > 40, > 65 and > 80 years of age in the general population. More favourable estimates of PPV were

observed as the prevalence of disease increases with age. The values obtained revealed 11 per cent PPV at 5.5 per cent prevalence, 20 per cent PPV at 10 per cent prevalence and 28 per cent PPV at 15 per cent prevalence in the general population.[41] The prevalence figures reflect conservative estimates of late-stage disease in the general population and would be further enhanced and more clinically applicable in a setting of diseased patients, as in the study conducted by Seddon and colleagues.[42] The longitudinal study design of the Age-Related Eye Disease Study (AREDS) cohort used in Seddon's study was ideal for evaluating incident AMD by distinguishing between 'progressors' and 'non-progressors' but, more importantly, it established that the same set of variants were effective at distinguishing non-disease controls from patients with late-stage disease. Not surprisingly, the same core panel of SNPs covering the major genes associated with disease used in Seddon and co-workers' test panel was also utilised in the study conducted by Jakobsdottir and colleagues,[41] as well as in our current study.

The present confirmatory findings reflect the utility of these variants to predict the development of CNV in non-diseased subjects in our study, as well as the progression to late-stage disease in patients diagnosed with early forms of AMD.[42] PPVs improve significantly when applied to the population of patients diagnosed with early stages of disease. The utility of AMD genetic testing will advance if the result of a predictive test translates into actionable information for the physician. This study highlights the need to continue to explore the biology of CNV, to improve our understanding of the genetics associated with disease and extend these findings in future studies to evaluate clinical performance metrics in the more acute clinical population diagnosed with early-stage disease. A genetic test identifying individuals at high risk of developing CNV holds the promise for earlier detection through risk-based surveillance protocols and improved outcomes arising from more timely intervention.

Acknowledgments

References

1. Rein, D.B., Wittenborn, J.S., Zhang, X. *et al.* (2009), 'Forecasting age-related macular degeneration through the year 2050: The potential impact of new treatment', *Arch. Ophthalmol.* Vol. 127, pp. 533–540.

2. Brown, G.C., Brown, M.M., Sharma, S. *et al.* (2006), 'The burden of age-related macular degeneration: A value-based medicine analysis', *Trans. Am. Ophthalmolol. Soc.* Vol. 103, pp. 173–186.

3. Klein, R., Klein, B.E.K., Jensen, S.C. and Meuer, S.M. (1997), 'The 5-year incidence and progression of age-related maculopathy. The Beaver Dam Eye Study', *Ophthalmology* Vol. 104, pp. 7–21.

4. Klein, R., Peto, T., Bird, A. and Vannewkirk, M.R. (2004), 'The epidemiology of age-related macular degeneration', *Am. J. Ophthalmol.* Vol. 137, pp. 486–495.

5. Allikmets, R. and Dean, M. (2008), 'Bringing age-related macular degeneration into focus', *Nat. Genet.* Vol. 40, pp. 820–821.

6. Swaroop, A., Chew, E.Y., Rickman, C.B. and Abecasis, G.R. (2009), 'Unraveling a multifactorial late-onset disease: From genetic susceptibility to disease mechanisms for age-related macular degeneration', *Annu. Rev. Genomics Hum. Genet.* Vol. 10, pp. 19–43.

7. Connell, P.P., Keane, P.A., O'Neill, E.C. *et al.* (2009), 'Risk factors for age-related maculopathy', *J. Ophthalmol.* Vol. 2009, Article ID 360764.

8. Scott, W.K., Schmidt, S., Hauser, M.A. *et al.* (2007), 'Independent effects of complement factor H Y402H polymorphism and cigarette smoking on risk of age-related macular degeneration', *Ophthalmology* Vol. 114, pp. 1151–1156.

9. DeAngelis, M.M., Ji, F, Kim, I.K. *et al.* (2007), 'Cigarette smoking, CFH, APOE, ELOVL4, and risk of neovascular age-related macular degeneration', *Arch. Ophthalmol.* Vol. 125, pp. 49–54.

10. DeAngelis, M.M., Ji, F, Adams, S. *et al.* (2008), 'Alleles in the HtrA serine peptidase 1 gene alter the risk of neovascular age-related macular degeneration', *Ophthalmology* Vol. 115, pp. 1209–1215.e7.

11. Dewan, A., Liu, M., Hartman, S. *et al.* (2006), 'HTRA1 promoter polymorphism in wet age-related macular degeneration', *Science* Vol. 314, pp. 989–992.

12. Edwards, A.O., Ritter, R., 3rd, Abel, K.J. *et al.* (2005), 'Complement factor H polymorphism and age-related macular degeneration', *Science* Vol. 308, pp. 421–424.

13. Hageman, G.S., Anderson, D.H., Johnson, L.V. *et al.* (2005), 'A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration', *Proc. Natl. Acad. Sci. USA* Vol. 102, pp. 7227–7232.

14. Haines, J.L., Hauser, M.A., Schmidt, S. *et al.* (2005), 'Complement factor H variant increases the risk of age-related macular degeneration', *Science* Vol. 308, pp. 419–421.

15. Jakobsdottir, J., Conley, Y.P., Weeks, D.E. *et al.* (2005), 'Susceptibility genes for age-related maculopathy on chromosome 10q26', *Am. J. Hum. Genet.* Vol. 77, pp. 389–407.

16. Klein, R.J., Zeiss, C., Chew, E.Y. *et al.* (2005), 'Complement factor H polymorphism in age-related macular degeneration', *Science* Vol. 308, pp. 385–389.

17. Li, M., Atmaca-Sonmez, P., Othman, M. *et al.* (2006), 'CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration', *Nat. Genet.* Vol. 38, pp. 1049–1054.

18. Rivera, A., Fisher, S.A., Fritsche, L.G. *et al.* (2005), '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.* Vol. 14, pp. 3227–3236.

19. Yang, Z., Camp, N.J., Sun, H. *et al.* (2006), 'A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration', *Science* Vol. 314, pp. 992–993.

20. Anderson, D.H., Mullins, R.F., Hageman, G.S. and Johnson, L.V. (2002), 'A role for local inflammation in the formation of drusen in the aging eye', *Am. J. Ophthalmol.* Vol. 134, pp. 411–431.

21. Anderson, D.H., Radeke, M.J., Gallo, N.B. *et al.* (2010), 'The pivotal role of the complement system in aging and age-related macular degeneration: hypothesis re-visited', *Prog. Retin. Eye Res.* Vol. 29, pp. 95–112.

22. Gehrs, K.M., Anderson, D.H., Johnson, L.V. and Hageman, G.S. (2006), 'Age-related macular degeneration — Emerging pathogenetic and therapeutic concepts', *Ann. Med.* Vol. 38, pp. 450–471.

23. Gehrs, K.M., Jackson, J.R., Brown, E.N. *et al.* (2010), 'Complement, age-related macular degeneration and a vision for the future', *Arch. Ophthalmol.* Vol. 128, pp. 349–358.

24. Hageman, G.S., Mullins, R.F., Russell, S.R. *et al.* (1999), 'Vitronectin is a constituent of ocular drusen and the vitronectin gene is expressed in human retinal pigmented epithelial cells', *FASEB J.* Vol. 13, pp. 477–484.

25. Hageman, G.S., Luthert, P.J., Victor Chong, N.H. *et al.* (2001), 'An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration', *Prog. Retin. Eye Res.* Vol. 20, pp. 705–732.

26. Mullins, R.F., Russell, S.R., Anderson, D.H. and Hageman, G.S. (2000), 'Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease', *FASEB J.* Vol. 14, pp. 835–846.

27. Ennis, S., Gibson, J., Cree, A.J. *et al.* (2010), 'Support for the involvement of complement factor I in age-related macular degeneration', *Eur. J. Hum. Genet.* Vol. 18, pp. 15–16.

28. Fagerness, J.A., Maller, J.B., Neale, B.M. *et al.* (2009), 'Variation near complement factor I is associated with risk of advanced AMD', *Eur. J. Hum. Genet.* Vol. 17, pp. 100–104.

29. Gold, B., Merriam, J.E., Zernant, J. *et al.* (2006), 'Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration', *Nat. Genet.* Vol. 38, pp. 458–462.

30. Hageman, G.S., Hancox, L.S., Taiber, A.J. *et al.* (2006), 'Extended haplotypes in the complement factor H (CFH) and CFH-related (CFHR) family of genes protect against age-related macular degeneration: Characterization, ethnic distribution and evolutionary implications', *Ann. Med.* Vol. 38, pp. 592–604.

31. Hughes, A.E., Orr, N., Esfandiary, H. *et al.* (2006), 'A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration', *Nat. Genet.* Vol. 38, pp. 1173–1177.

32. Jakobsdottir, J., Conley, Y.P., Weeks, D.E. *et al.* (2008), 'C2 and CFB genes in age-related maculopathy and joint action with CFH and LOC387715 genes', *PLoS. ONE* Vol. 3, p. e2199.

33. Kondo, N., Bessho, H., Honda, S. and Negi, A. (2010), 'Additional evidence to support the role of a common variant near the complement factor I gene in susceptibility to age-related macular degeneration', *Eur. J. Hum. Genet.* Vol. 18, pp. 634–635.

34. Maller, J., George, S., Purcell, S. *et al.* (2006), 'Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration', *Nat. Genet.* Vol. 38, pp. 1055–1059.

35. Maller, J.B., Fagerness, J.A., Reynolds, R.C. *et al.* (2007), 'Variation in complement factor 3 is associated with risk of age-related macular degeneration', *Nat. Genet.* Vol. 39, pp. 1200–1201.

36. Spencer, K.L., Olson, L.M., Anderson, B.M. *et al.* (2008), 'C3 R102G polymorphism increases risk of age-related macular degeneration', *Hum. Mol. Genet.* Vol. 17, pp. 1821–1824.

37. Yates, J.R., Sepp, T., Matharu, B.K. *et al.* (2007), 'Complement C3 variant and the risk of age-related macular degeneration', *N. Engl. J. Med.* Vol. 357, pp. 553–561.

38. Zhang, H., Morrison, M.A., Dewan, A. *et al.* (2008), 'The NEI/NCBI dbGAP database: Genotypes and haplotypes that may specifically predispose to risk of neovascular age-related macular degeneration', *BMC Med. Genet.* Vol. 9, p. 51.

39. Chen, W., Stambolian, D., Edwards, A.O. *et al.* (2010), 'Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration', *Proc. Natl. Acad. Sci. USA* Vol. 107, pp. 7401–7406.

40. Neale, B.M., Fagerness, J., Reynolds, R. *et al.* (2010), 'Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC)', *Proc. Natl. Acad. Sci. USA* Vol. 107, pp. 7395–7400.

41. Jakobsdottir, J., Gorin, M.B., Conley, Y.P. *et al.* (2009), 'Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers', *PLoS Genet.* Vol. 5, p. e1000337.

42. Seddon, J.M., Reynolds, R., Maller, J. *et al.* (2009), 'Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables', *Invest. Ophthalmol. Vis. Sci.* Vol. 50, pp. 2044–2053.

43. McKay, G.J., Dasari, S., Patterson, C.C. *et al.* (2010), 'Complement component 3: An assessment of association with AMD and analysis of gene-gene and gene-environment interactions in a Northern Irish cohort', *Mol. Vis.* Vol. 16, pp. 194–199.

44. Zanke, B., Hawken, S., Carter, R. and Chow, D. (2010), 'A genetic approach to stratification of risk for age-related macular degeneration', *Can. J. Ophthalmol.* Vol. 45, pp. 22–27.

45. Brown, E.N., Hancox, L.S., Miller, N.J. *et al.* (2010), 'Determination and assessment of extended haplotypes spanning the chromosome 1q32 CFH-to-CFHR5 locus', *Invest. Ophthalmol. Vis. Sci.* Vol. 51, Abstract 1262.

46. Baird, P.N., Islam, F.M., Richardson, A.J. *et al.* (2006), 'Analysis of the Y402H variant of the complement factor H gene in age-related macular degeneration', *Invest. Ophthalmol. Vis. Sci.* Vol. 47, pp. 4194–4198.

47. Robman, L., Baird, P.N., Dimitrov, P.N. *et al.* (2010), 'C-reactive protein levels and complement factor H polymorphism interaction in age-related macular degeneration and its progression', *Ophthalmology* Vol. 117, pp. 1982–1988.

48. National Center for Biotechnology Information, National Library of Medicine Database of Single Nucleotide Polymorphisms (dbSNP). http://www.ncbi.nlm.nih.gov/SNP. (Accessed 4th April, 2011).

49. Kent, W.J., Sugnet, C.W., Furey, T.S. *et al.* (2002), 'The human genome browser at UCSC', *Genome Res.* Vol. 12, pp. 996–1006.

50. Ambroise, C. and McLachlan, G. (2002), 'Selection bias in gene extraction on the basis of microarray gene-expression data', *Proc. Natl. Acad. Sci. USA* Vol. 99, pp. 6562–6566.

51. DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988), 'Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach', *Biometrics* Vol. 44, pp. 837–845.

52. SAS Institute Inc. (2008), 'SAS Version 9.1', SAS Institute, Cary, NC, USA.

53. Rosner, B. and Glynn, R.J. (2009), 'Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of C statistics from alternative prediction models', *Biometrics* Vol. 65, pp. 188–197.

Supplementary Analysis 1. Logistic regression results

Model information

Dataset	WORK.SORT8168
Response variable	Response
Number of response levels	2
Model	Binary logit
Optimisation technique	Fisher's scoring
Number of observations read	1,000
Number of observations used	949

Response profile

Ordered value	Response	Total frequency
1	0	467
2	1	482

Probability modelled is response = 0.
Note: 51 observations were deleted due to missing values for the response or explanatory variables.

Backward elimination procedure

Step 0. The following effects were entered:

Intercept rs10490924 rs1061170 rs10922153 rs12144939 rs1409153 rs1750311 rs2230199 rs2274700 rs2990510 rs403846 rs641153 rs698859 rs9332739

Model convergence status

Convergence criterion (GCONV = 1E-8) satisfied.

Model fit statistics

Criterion	Intercept only	Intercept and covariates
AIC	1317.356	1016.228
SC	1322.212	1084.204
−2 Log L	1315.356	988.228

Testing global null hypothesis: BETA = 0

Test	Chi-square	DF	Pr > ChiSq
Likelihood ratio	327.1278	13	<0.0001

Continued

Continued

Testing global null hypothesis: BETA = 0

Test	Chi-square	DF	Pr > ChiSq
Score	280.8660	13	<0.0001
Wald	209.1689	13	<0.0001

Step 1. Effect rs698859 is removed:

Model convergence status

Convergence criterion (GCONV = 1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept only	Intercept and covariates
AIC	1317.356	1014.231
SC	1322.212	1077.352
−2 Log L	1315.356	988.231

Testing global null hypothesis: BETA = 0

Test	Chi-square	DF	Pr > ChiSq
Likelihood ratio	327.1248	12	<0.0001
Score	280.8660	12	<0.0001
Wald	209.1627	12	<0.0001

Residual Chi-square test

Chi-Square	DF	Pr > ChiSq
0.0031	1	0.9559

Step 2. Effect rs1409153 is removed:

Model convergence status

Convergence criterion (GCONV = 1E-8) satisfied.

Model fit statistics

Criterion	Intercept only	Intercept and covariates
AIC	1317.356	1012.567
SC	1322.212	1070.832
−2 Log L	1315.356	988.567

Testing global null hypothesis: BETA = 0 Test	Chi-square	DF	Pr > ChiSq
Likelihood ratio	326.7893	11	<0.0001
Score	280.6633	11	<0.0001
Wald	209.0053	11	<0.0001

Residual Chi-square test Chi-square		DF	Pr > ChiSq
0.3389		2	0.8441

Step 3. Effect rs1750311 is removed:

Model convergence status
Convergence criterion (GCONV = 1E-8) satisfied.

Model fit statistics Criterion	Intercept only	Intercept and covariates
AIC	1317.356	1010.949
SC	1322.212	1064.358
−2 Log L	1315.356	988.949

Testing global null hypothesis: BETA = 0 Test	Chi-square	DF	Pr > ChiSq
Likelihood ratio	326.4077	10	<0.0001
Score	280.4794	10	<0.0001
Wald	209.1743	10	<0.0001

Residual Chi-square test Chi-Square		DF	Pr > ChiSq
0.7200		3	0.8685

Step 4. Effect rs12144939 is removed:

Model convergence status
Convergence criterion (GCONV = 1E-8) satisfied.

Model fit statistics Criterion	Intercept only	Intercept and covariates
AIC	1317.356	1010.903
SC	1322.212	1059.457
−2 Log L	1315.356	990.903

Testing global null hypothesis: BETA = 0 Test	Chi-square	DF	Pr > ChiSq
Likelihood ratio	324.4536	9	<0.0001
Score	279.2738	9	<0.0001
Wald	209.2428	9	<0.0001

Residual Chi-square test Chi-square		DF	Pr > ChiSq
2.6773		4	0.6132

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of backward elimination					
Step	Effect removed	DF	Number in	Wald Chi-square	Pr > ChiSq
1	rs698859	1	12	0.0031	0.9559
2	rs1409153	1	11	0.3356	0.5624
3	rs1750311	1	10	0.3820	0.5366
4	rs12144939	1	9	1.9468	0.1629

Analysis of maximum likelihood estimates					
Parameter	DF	Estimate	Standard error	Wald Chi-square	Pr > ChiSq
Intercept	1	−0.7554	0.2621	8.3051	0.0040
rs10490924	1	−1.4417	0.1245	134.0342	<0.0001
rs1061170	1	0.7697	0.2988	6.6352	0.0100
rs10922153	1	0.7240	0.1950	13.7839	0.0002
rs2230199	1	−0.4292	0.1286	11.1389	0.0008
rs2274700	1	0.8593	0.1695	25.7009	<0.0001

Continued

Continued

Analysis of maximum likelihood estimates

Parameter	DF	Estimate	Standard error	Wald Chi-square	Pr > ChiSq
rs2990510	1	0.4556	0.1586	8.2557	0.0041
rs403846	1	−0.6775	0.3341	4.1118	0.0426
rs641153	1	0.8243	0.1999	17.0040	<0.0001
rs9332739	1	0.9509	0.3163	9.0360	0.0026

Odds ratio estimates

Effect	Point estimate	95% Wald confidence limits	
rs10490924	0.237	0.185	0.302
rs1061170	2.159	1.202	3.878
rs10922153	2.063	1.407	3.023
rs2230199	0.651	0.506	0.838

Continued

Continued

Odds ratio estimates

Effect	Point estimate	95% Wald confidence limits	
rs2274700	2.362	1.694	3.292
rs2990510	1.577	1.156	2.152
rs403846	0.508	0.264	0.978
rs641153	2.280	1.541	3.374
rs9332739	2.588	1.392	4.811

Association of predicted probabilities and observed responses

Percentage concordant	81.5	**Somers' D**	0.637
Percentage discordant	17.9	**Gamma**	0.641
Percentage tied	0.6	**Tau-a**	0.319
Pairs	225094	**c**	0.818

Classification table

Prob. Level	Correct Event	Correct Non-event	Incorrect Event	Incorrect Non-event	Percentages Correct	Sensitivity	Specificity	False positive	False negative
0.000	467	0	482	0	49.2	100.0	0.0	50.8	—
0.020	467	3	479	0	49.5	100.0	0.6	50.6	0.0
0.040	467	20	462	0	51.3	100.0	4.1	49.7	0.0
0.060	467	35	447	0	52.9	100.0	7.3	48.9	0.0
0.080	467	49	433	0	54.4	100.0	10.2	48.1	0.0
0.100	465	65	417	2	55.8	99.6	13.5	47.3	3.0
0.120	461	91	391	6	58.2	98.7	18.9	45.9	6.2
0.140	457	113	369	10	60.1	97.9	23.4	44.7	8.1
0.160	450	143	339	17	62.5	96.4	29.7	43.0	10.6
0.180	448	159	323	19	64.0	95.9	33.0	41.9	10.7

Continued

Continued

Classification table

Prob. Level	Correct Event	Correct Non-event	Incorrect Event	Incorrect Non-event	Percentages Correct	Sensitivity	Specificity	False positive	False negative
0.200	442	182	300	25	65.8	94.6	37.8	40.4	12.1
0.220	438	200	282	29	67.2	93.8	41.5	39.2	12.7
0.240	435	213	269	32	68.3	93.1	44.2	38.2	13.1
0.260	434	217	265	33	68.6	92.9	45.0	37.9	13.2
0.280	423	227	255	44	68.5	90.6	47.1	37.6	16.2
0.300	422	246	236	45	70.4	90.4	51.0	35.9	15.5
0.320	419	252	230	48	70.7	89.7	52.3	35.4	16.0
0.340	414	271	211	53	72.2	88.7	56.2	33.8	16.4
0.360	410	274	208	57	72.1	87.8	56.8	33.7	17.2
0.380	389	287	195	78	71.2	83.3	59.5	33.4	21.4
0.400	385	303	179	82	72.5	82.4	62.9	31.7	21.3
0.420	381	312	170	86	73.0	81.6	64.7	30.9	21.6
0.440	365	326	156	102	72.8	78.2	67.6	29.9	23.8
0.460	361	331	151	106	72.9	77.3	68.7	29.5	24.3
0.480	358	340	142	109	73.6	76.7	70.5	28.4	24.3
0.500	344	354	128	123	73.6	73.7	73.4	27.1	25.8
0.520	332	357	125	135	72.6	71.1	74.1	27.4	27.4
0.540	324	366	116	143	72.7	69.4	75.9	26.4	28.1
0.560	315	378	104	152	73.0	67.5	78.4	24.8	28.7
0.580	300	389	93	167	72.6	64.2	80.7	23.7	30.0
0.600	293	392	90	174	72.2	62.7	81.3	23.5	30.7
0.620	284	398	84	183	71.9	60.8	82.6	22.8	31.5
0.640	266	410	72	201	71.2	57.0	85.1	21.3	32.9
0.660	252	417	65	215	70.5	54.0	86.5	20.5	34.0
0.680	236	423	59	231	69.4	50.5	87.8	20.0	35.3
0.700	233	427	55	234	69.5	49.9	88.6	19.1	35.4
0.720	196	440	42	271	67.0	42.0	91.3	17.6	38.1
0.740	190	441	41	277	66.5	40.7	91.5	17.7	38.6
0.760	179	448	34	288	66.1	38.3	92.9	16.0	39.1

Continued

Continued

Classification table

| Prob. Level | Correct | | Incorrect | | Percentages | | | | |
	Event	Non-event	Event	Non-event	Correct	Sensitivity	Specificity	False positive	False negative
0.780	170	453	29	297	65.6	36.4	94.0	14.6	39.6
0.800	127	456	26	340	61.4	27.2	94.6	17.0	42.7
0.820	114	467	15	353	61.2	24.4	96.9	11.6	43.0
0.840	103	467	15	364	60.1	22.1	96.9	12.7	43.8
0.860	77	470	12	390	57.6	16.5	97.5	13.5	45.3
0.880	65	471	11	402	56.5	13.9	97.7	14.5	46.0
0.900	53	475	7	414	55.6	11.3	98.5	11.7	46.6
0.920	40	479	3	427	54.7	8.6	99.4	7.0	47.1
0.940	16	481	1	451	52.4	3.4	99.8	5.9	48.4
0.960	9	481	1	458	51.6	1.9	99.8	10.0	48.8
0.980	1	481	1	466	50.8	0.2	99.8	50.0	49.2
1.000	0	482	0	467	50.8	0.0	100.0	—	49.2