HAL
open science

Determination of a gene and environment risk model for age related macular degeneration.

Jane Gibson, Angela J Cree, Andrew Collins, Andrew Lotery, Sarah Ennis

To cite this version:

Jane Gibson, Angela J Cree, Andrew Collins, Andrew Lotery, Sarah Ennis. Determination of a gene and environment risk model for age related macular degeneration.. British Journal of Ophthalmology, 2010, 94 (10), pp.1382. 10.1136/bjo.2010.182568 . hal-00557364

HAL Id: hal-00557364

https://hal.science/hal-00557364v1

Submitted on 19 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Determination of a gene and environment risk model for age related macular degeneration.

Jane Gibson[1], Angela Cree[2], Andrew Collins[1], Andrew Lotery[2,3], Sarah Ennis[1*].

Keywords: Age-related Macular Degeneration, Association, *SERPING1*, risk model.

[1] University of Southampton, Genetic Epidemiology & Bioinformatics Group. Duthie Building (Mp 808). Southampton General Hospital, Tremona Road, Southampton. SO16 6YD, UK

[2] University of Southampton, Clinical Neurosciences Division (Mp 806), Southampton General Hospital, Southampton, SO16 6YD, UK

[3] Southampton Eye Unit, Southampton General Hospital, Southampton, SO16 6YD, UK

* Corresponding author: Dr. S. Ennis
Genetic Epidemiology & Bioinformatics Group
University of Southampton
Human Genetics Division (Mailpoint 808)
Southampton General Hospital,
Southampton,
SO16 6YD, United Kingdom
E-mail: se@soton.ac.uk
Tel: +44 (0)2380 798614 Fax: +44 (0)2380 794264

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in British Journal of Ophthalmology and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence (http://group.bmj.com/products/journals/instructions-for-authors/licence-forms).

Competing Interest: None to declare.

Abstract

Background/aims: Recently we identified association between age related macular degeneration (AMD) and genetic variants in the Serpin peptidase inhibitor, clade G, member 1 (*SERPING1*) gene. In the current study we interrogated the genomic region in linkage disequilibrium (LD) with the *SERPING1* gene, and modelled the contribution to disease of known genetic and environmental AMD risk factors.

Methods: We analysed genes neighbouring *SERPING1* and examined haplotype association with AMD. A stepwise logistic regression model was developed including known genetic and environmental risk factors (age, gender and smoking). Individual risk scores were assessed between groups of cases and controls.

Results: In the *SERPING1* region rs2511989 remains most significantly associated (p=1.77E-05, odds ratio of 0.67). One haplotype, containing the rs2511989 variant and the majority of *SERPING1*, exhibits marginally stronger association (p=5.13E-06, odds ratio of 0.66). Our risk model includes six SNPs in *CFH, C3, HTRA1* and *SERPING1*, showing independent effects, which together account for 45% of risk of developing AMD (p=1.65E-50) with a combined population attributable risk (PAR) of 87%.

Conclusion: Results implicate *SERPING1*, with no convincing evidence for involvement of other genes in the region. We demonstrate a multifactorial model with significant differences in risk scores for cases versus controls (p=9.81E-71) and across AREDS stratified cases (p=1.88E-11).

Introduction

Age-related macular degeneration (AMD) is the most common cause of visual loss in the developed world[1]. Several genes have been associated with the disease including complement factor H (*CFH*)[2], the age-related maculopathy susceptibility 2/HtrA serine peptidase 1 (*ARMS2/HTRA1*) region of chromosome 10q[3], complement component 3 (*C3*)[4], complement factor B (*CFB*) and complement component 2 (*C2*)[5]. Recently the *SERPING1* gene has also been associated with AMD and replicated in an independent US sample[6]. This finding was not replicated in a further 3 publications[7-9] but was replicated in a sample using UK controls[10]. Lu et al. also found association with a haplotype in the *SERPING1* gene and the phenotype soft drusen[11].

Ennis *et al.* analysed genetic variation in the form of single nucleotide polymorphisms (SNPs) within and immediately adjacent to the *SERPING1* gene. This approach could not exclude the possibility that other nearby variants generally inherited together on the same chromosomal background (in linkage disequilibrium, LD) as the *SERPING1* locus accounted for the association with disease[6]. Global gene expression data[12] (*mRNA by SNP browser* v1.0.1, http://www.sph.umich.edu/csg/liang/asthma/) also indicate that expression of the *TIMM10* gene located 67kilobases (kb) proximal to *SERPING1*, is highly associated (p=1.50E-21) with SNPs less than 9kb from *SERPING1*. The current study extends the region previously studied to encompass the chromosomal region, including the *TIMM10* gene, in strong linkage disequilibrium with the *SERPING1* gene. We aim to test if association between AMD and other variants in the region can explain the observed association between *SERPING1* and AMD.

A number of environmental factors also affect AMD predisposition including age, ethnicity, family history, gender, and cigarette smoking[13]. Nutritional factors such as omega-3 intake have also been show to modify disease risk[14]. Several analyses of multiple factors affecting AMD risk

and progression have been conducted. Francis *et al*.[15] analysed the joint effects of *CFH* and the chromosome 10q locus and concluded both loci contributed independent effects. Seitsonen *et al.*[16] analysed variants in *CFH, ARMS2* and *C3* determining a population attributable risk (PAR – *see methods for definition*) of ~65% for the 3 risk genes. They also found smoking had an effect only when other variables such as gender and the *C3* genotype were accounted for. Seddon *et al.*[17] analysed 6 SNPs in *CFH, ARMS2, C2, CFB* and *C3* as well as non-genetic factors such as smoking, BMI, and those taking antioxidants as part of a clinical trial. They found independent effects of these polymorphisms and smoking on prevalence of AMD and developed a prediction model for progression and incidence of AMD. All of these analyses pre-dated the identification of the *SERPING1* locus in disease susceptibility. Our present study creates a multifactorial model of risk which will assess the effects of known genes including *SERPING1* as well as environmental effects. We determine the degree to which variants in the *SERPING1* region act independently from established risk factors and estimate the overall risk accounted for by the model. These data enable calculation of a score for each individual showing their risk of AMD, these 'risk scores' are compared across groups.

Materials and Methods

Cohort;

Samples used in this study overlapped with the samples from the UK cohort previously described in Ennis et al. 2008[6] with some samples being replaced due to diminished DNA stocks. Recruitment was approved by the Southampton and Southwest Hants local research ethics committee and followed the tenets of the Declaration of Helsinki. All participants provided informed written consent. The sample included 470 cases and 470 controls, and consisted of 544 females and 396 males. Clinically, AMD can be classified with a grading system as used in the Age Related Eye Disease Study (AREDS)[18]. For the present study these classes were defined as follows: an AREDS score of two (69 cases) was defined as multiple small drusen,

single or nonextensive intermediate drusen [63-124 μm], pigment abnormalities, or any combination of these; an AREDS score of three (116 cases) included the absence of advanced AMD and at least 1 large druse (125 μm), extensive intermediate drusen, or geographic atrophy (GA) that did not involve the center of the macula, or any combination of these and an AREDS score of four (285 cases) was classified as lesions of advanced AMD in the central macula (including the fovea) which included the presence of either choroidal neovascularization, disciform macular scarring or geographic atrophy.

SNP selection;

A region around the *SERPING1* gene on chromosome 11q12.1 was defined using a linkage disequilibrium unit (LDU) map[19] of publicly available genetic data on a sample with European ancestry (www.hapmap.org). This region extends from one LDU 5' of *SERPING1* to one LDU 3' of the gene spanning 550 kb (56,941,540 to 57,491,647bp:NCBI build 36.1) and shows the region including variants likely to be inherited together (Figure 1). Haplotype tagging SNPs, were selected to reduce redundancy in the data, since only one of a pair or group of SNPs giving the same information is selected (Tagger Pairwise, r^2 <0.9). All HapMap SNPs in the *SERPING1* gene were genotyped even though some are redundant (LD measure, r^2>0.8). Both HapMap SNPs in *TIMM10* gene (rs17453436, rs10792101) were included. Quality control procedures removed very rare SNPs (minor allele frequency <5%) and those with genotype counts which deviated largely from expected indicating genotyping error (Hardy-Weinberg equilibrium in the control sample of p< 0.001). A total of 61 SNPs were analysed in the SERPING1 region including 11 SNPs from the original study[6] (Supplementary Table 1).

We also genotyped 35 SNPs in 20 genes previously associated with AMD. After quality control, 26 SNPs in 15 genes were analysed. Genotyping was carried out using KASPar chemistry (http://www.kbioscience.co.uk/genotyping/genotyping_chemistry.html).

SNP and haplotype analysis;

Each SNP was analysed by a chi square of the allele counts in cases and controls, associated p-values and odds ratios with confidence intervals were calculated. Chi squares were also calculated for other genetic models including the genotypic test, Cochran-Armitage trend test, and recessive and dominant models using PLINK[20]. We analysed SNPs together in haplotypes within 10 smaller regions (blocks) where the SNPs are more commonly inherited together (defined on a solid spine of LD with a D'=0.8 by HAPLOVIEW, Supplementary Figure 1). We determined the possible combinations of SNPs (estimated haplotypes) present at >1% frequency over the *SERPING1* region[21] and tested for association with AMD risk. Omnibus haplotype analysis detects a unique effect across all haplotypes in a block. Haplotype specific association tests each haplotype versus the rest in each block. Multiple testing correction was carried out by Bonferroni or permutation.

Logistic regression model;

To determine the contribution of the *SERPING1* gene to AMD risk in the context of other risk genes we conducted logistic regression analysis. This accounts for the variation in case/control status (dependent variable) using the best combination of independent variables, including any Bonferroni corrected significant SNPs from the *SERPING1* region, the 26 SNPs from known AMD risk genes, age, gender, and smoking status. SNPs were coded, for example, aa=0, aA=1 and AA=2.

Population attributable risk (PAR);

PAR measures the effect on disease rate in the population that can be attributed to a particular factor, showing the percentage of cases that could be avoided if the risk factor were eradicated. This is generally not feasible, for example, in the case of genetic risk factors or gender, but is a

useful measure of the importance of each risk factor. We calculated this as, PAR=Pr(OR-1)/(1+Pr(OR-1)), where OR is the odds ratio from the logistic regression model and Pr the prevalence of the risk factor in the general population, estimated from controls[16]. A combined PAR was also calculated as $PAR=1-(1-PAR_1)(1-PAR_2)\ldots(1-PAR_n)$ incorporating all risk factors.

Individuals risk scores;

Using variable coefficients (B) from the logistic regression, which show the effect size of each risk factor (natural log of the odds ratio), we calculated a risk score for each individual using previously described methods[17]. The risk score is a value for each individual intended to convey their personal risk of developing AMD. Risk scores were then compared between known cases and controls by an ANOVA test. Receiver operating characteristic (ROC) curve analysis determined the classification accuracy of the risk scores by calculating the area under the curve (AUC) with 95% confidence intervals and significance level. We further investigated the risk scores in the patient group stratified by AREDS score and carried out an ANOVA test followed by a *post hoc* least significant difference (LSD) test.

All analyses were carried out using PLINK v1.04[20], HAPLOVIEW v4.1[21] and SPSS(v16.0).

Results

Single SNP and haplotype analysis;

There were several results with uncorrected P-values less than 0.05 across the *SERPING1* region, as expected given extended LD. Figure 1 shows a cluster of significant results which corresponding to the *SERPING1* gene. Five SNPs (rs2511989, rs2509897, rs11229109, rs2244169 and rs2511990) have a $p<0.05$ after multiple testing correction (Table 1). The most significant result is for the *SERPING1* SNP rs2511989 (p=1.10E-03). The protective effect is shown by an odds ratio (OR) of 0.67 (95% CI 0.56-0.80). Similar results for alternative genetic

models are given in Supplementary Table 3. SNPs within and adjacent to the *TIMM10* gene show no association (p=0.61 and 0.76).

Omnibus haplotype analysis (after Bonferroni correction) gives significant results in blocks six and seven (Table 3). Block six is upstream of the *SERPING1* gene and contains the promoter region including the second most significant SNP rs2509897 (p=1.60E-03) six haplotypes within this block account for 97.4% of our sample. Block seven includes all SNPs genotyped within the *SERPING1* gene and contains three haplotypes accounting for 98.1% of the sample. The common *GCTCGTTGGG* haplotype in block seven demonstrates the strongest association and contains rs2511989, the SNP most strongly associated with AMD risk. Although the significance of the haplotype is marginally greater than that observed for the single SNP rs2511989, the odds ratios are almost identical (0.65 and 0.67 respectively).

Single SNP allelic chi square results for variants implicated in AMD risk, in genes other than *SERPING1*, are presented in Supplementary table 2. SNP rs10490924 within *ARMS2* showed the strongest association (p=3.65E-20). SNPs in *CFH, HTRA1, CFHR5, C3, APOE, C2* and *C7* were also significant before correction. Ten SNPs in *CFH, CHFR5, HTRA1, ARMS2* and *C3* remain significant after correction for multiple testing.

Logistic regression model and population attributable risk (PAR) estimates;

The five SNPs significant after correction from the *SERPING1* region (Table 1), all 26 variants from other genomic regions as well as age, gender and smoking status were analysed by logistic regression to determine their contribution to AMD disease status. Six SNPs were significant (Table 4). Similar to previous analyses, inclusion of multiple SNPs from the *CFH* gene shows multiple independent effects in this gene. The *HTRA1, SERPING1* and *C3* loci were also represented. Of the non-genetic factors, advanced age, female gender and positive

smoking status also contribute to AMD status. The model shows the proportion of variation in risk accounted for as the Nagelkerke R^2= 0.446 and a good overall fit (p=1.65E-50). Despite the similarity in odds ratios for *SERPING1* and *C3* (although *C3* shows a risk effect and *SERPING1* a protective effect) their respective PAR values differ reflecting the difference in risk allele prevalence (0.19 in *C3* and 0.56 in *SERPING1*). The combined population attributable risk (PAR) based on these data is 87% (Table 4).

Individuals risk scores;

Risk scores (Figure 2, A and B) show a highly significant (p= 9.81E-71) shift between cases (mean score= 0.81) and controls (mean score= -1.09), however, there is appreciable overlap. A high area under the curve (AUC) value of 0.83 (95% CI 0.81-0.86) was obtained by ROC curve analysis as a measure of classification accuracy, showing an 83% chance of correctly assigning a random case and a random control correctly. The point on the curve farthest from the diagonal line of chance shows the optimal threshold for assigning cases and controls with a 76% probability of correctly assigning a case and a 24% probability of misclassifying a control as a case (Supplementary Figure 2). We stratified the patient group by AREDS score (Figure 2, C, D and E) and observed a significant difference in risk score across groups (p=1.88E-11). A *post hoc* least significant difference (LSD) test shows the difference lies between cases with an AREDS score of two versus either group of cases with scores of AREDS three or four. The risk scores for the group containing cases with an AREDS score of two versus the group with an AREDS score of 4 were significantly different (p=4.94E-12), as were the risk scores in the groups containing cases with AREDS scores of two versus three (p=4.07E-09).

Discussion

Age-related macular degeneration (AMD) was one of the first complex diseases to be successfully analysed by genome-wide association. Since the initial successful identification of

the *CFH* gene, several other genes have been implicated and replicated. Recently the *SERPING1* gene was added to this list after a highly significant association signal was observed as part of a candidate gene study and replicated in an independent US sample[6]. There have been mixed results in further studies, several have failed to replicate this result[7-9] while others have found positive results[10;11].

The current study aimed to distinguish if the association signals implicating the *SERPING1* locus might be acting in proxy for some other linked variant(s). In particular, the *TIMM10* gene has previously shown altered expression levels associated with genetic variants close to the *SERPING1* promoter[12]. However, SNPs within and immediately adjacent to *TIMM10* showed no association with AMD. Across the region analysed, rs2511989 located within intron six of the *SERPING1* gene remains the most significantly associated with AMD as previously reported[6], a number of markers including rs2509897 in the promoter region are strongly supportive. The pattern of associated markers after correction for multiple testing shows clear clustering limited to the *SERPING1* gene itself. A single exception is for rs11229109 in the *TXNDC14* gene. This gene is widely expressed and postulated to play a role in monitoring protein folding in the endoplasmic reticulum[22]. Although *TXNDC14* involvement in AMD cannot be unequivocally ruled out, the association is less significant than that seen with *SERPING1* gene and *TXNDC14* is not a conspicuous functional candidate in the aetiology of macular degeneration. Whereas, the SERPING1 gene encodes the complement component 1 inhibitor, an important first step in the regulation of the complement system which has been highly implicated in AMD pathogenesis.

The *GCTCGTTGGG* haplotype within the *SERPING1* gene is most strongly associated with disease and marginally more significant than SNP rs2511989. rs2511989 is the only SNP present in this haplotype as a rare allele explaining why haplotype analysis does not give

appreciably more information than the single SNP test. SNP rs2511988 lies adjacent to rs2511989 within this haplotype and in high LD. Although nominally significant rs2511988 does not withstand correction for multiple testing (p=0.07). However, there is evidence suggesting that variation at rs2511988 may affect splicing efficiency and intron six removal since it is located directly before exon seven and adjacent to a branch point sequence[23]. Other significant SNPs are located in the promoter region of *SERPING1* suggesting possible regulatory effects.

We assessed the contribution of the *SERPING1* locus in the context of other genetic and environmental factors. Our multifactorial model showed that only variants from the *CFH, HTRA1, C3* and *SERPING1* genes showed sustained effects. This is the first study to show the independent effects of *SERPING1* given the effects of other well established genes. Interestingly, for a given gene, the SNP variants which contributed to the final model were not necessarily the most significantly associated SNPs when assessed independently. For example, the *SERPING1* promoter SNP rs2509897 was included over rs2511989. Increasing age, female gender and smoking also significantly contributed to our model which accounts for a high proportion, 45%, of AMD risk.

Population attributable risks (PARs) were calculated for each significant risk factor in our model. Our data shows the *CFH* variant rs1061170 has a PAR of 35%, slightly lower than in previous publications, 49%[24] and 58.1%[16]. The *C3* rs2230199 variant has a PAR of ~8% in our data similar to published results, 17%[24] and 5.8%[16]. In comparison with estimates for other risk genes a PAR of 17% for the *SERPING1* variant rs2509897 indicates the importance of this gene in our population. The combined PAR of 87% in our data may imply that determining an individual's risk of disease based on these factors alone would be feasible. Accurate disease prediction is clinically valuable in terms of early diagnosis, and modification of risk factors such as smoking and nutrition or with earlier drug intervention. Our results show a significant

difference in risk scores between cases and controls and in cases grouped by ARED score (Figure 2) and ROC curve analysis shows the discrimination power of this model is 83% (Supplementary Figure 2). This shows there is an 83% chance of predicting a random case and a random control correctly. Our analysis shows that if 76% of cases were correctly classified, 24% of controls would be misclassified as cases. A high risk score may justify more regular ophthalmic examinations for those who do not appear to have AMD. However, given a disease prevalence of ~5%, of a 1000 people, 228 (86%) of the 266 considered high risk would be wrongly assigned and a substantial proportion of cases would be missed. Therefore, although our risk prediction model is supported by a high discrimination power, accuracy is not sufficient for direct applicability in a clinical setting.

Our model may be limited by noise in the data, for example controls may yet develop this age-related disease and AMD cases may progress to a more advanced form. We have limited this noise by selecting controls over the age of 50 years. Although all major AMD risk genes were included, incomplete knowledge of the risk factors affecting AMD may limit this study and inclusion of new risk factors should improve the resolution of the model. The odds ratios we present are broadly comparable to those published previously suggesting this model would be generally applicable; however, it should be tested in an independent cohort of patients.

We believe the value of this study is that it clarifies the contribution of risk factors which lead to the development of AMD. However, it also highlights that such determination does not guarantee effective discrimination between cases and controls. Others have found similar results[25]. This is an important result in an era where increasingly commercial testing is being directly offered to patients to determine their genetic risk of AMD (http://www.decodeme.com , https://www.23andme.com). The future of personalised medicine promises the use of genetic

testing and risk prediction for early diagnosis and treatment, but our results show that we should avoid overstating association findings at an individual level.

References

1. Friedman DS, O'Colmain BJ, Munoz B *et al*. Prevalence of age-related macular degeneration in the United States. *Arch.Ophthalmol.* 2004;**122**:564-72.

2. Klein RJ, Zeiss C, Chew EY *et al*. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;**308**:385-9.

3. Jakobsdottir J, Conley YP, Weeks DE *et al*. Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am.J.Hum.Genet.* 2005;**77**:389-407.

4. Yates JR, Sepp T, Matharu BK *et al*. Complement C3 variant and the risk of age-related macular degeneration. *N.Engl.J.Med.* 2007;**357**:553-61.

5. Spencer KL, Hauser MA, Olson LM *et al*. Protective effect of complement factor B and complement component 2 variants in age-related macular degeneration. *Hum.Mol.Genet.* 2007;**16**:1986-92.

6. Ennis S, Jomary C, Mullins R *et al*. Association between the SERPING1 gene and age-related macular degeneration: a two-stage case-control study. *Lancet* 2008;**372**:1828-34.

7. Park KH, Ryu E, Tosakulwong N *et al*. Common variation in the SERPING1 gene is not associated with age-related macular degeneration in two independent groups of subjects. *Mol.Vis.* 2009;**15**:200-7.

8. Carter JG, Churchill AJ. Analysis of SERPING1 and its association with age-related macular degeneration. *Acta Ophthalmol.* 2009.

9. Allikmets R, Dean M, Hageman GS *et al*. The SERPING1 gene and age-related macular degeneration. *Lancet* 2009;**374**:875-6.

10. Ramsden C, Cooper M, Tan A *et al*. SERPING1 Mutation rs2511989 is Variably Associated with Macular Degeneration and Not Associated with Severity of Disease. (Abstract #1911). *Presented at the 59th Annual Meeting of The American Society of Human Genetics, October 22, 2009, Honolulu, Hawaii.* 2009;**http://www.ashg.org/2009meeting**.

11. Lu F, Zhao P, Fan Y *et al*. An association study of SERPING1 gene and age-related macular degeneration in a Han Chinese population. *Molecular Vision* 2010;**16**:1-6.

12. Dixon AL, Liang L, Moffatt MF *et al*. A genome-wide association study of global gene expression. *Nat.Genet.* 2007;**39**:1202-7.

13. Age-Related Eye Disease Study Research Group. Risk factors associated with age-related macular degeneration. A case-control study in the age-related eye disease study: Age-Related Eye Disease Study Report Number 3. *Ophthalmology* 2000;**107**:2224-32.

14. Seddon JM, George S, Rosner B. Cigarette smoking, fish consumption, omega-3 fatty acid intake, and associations with age-related macular degeneration: the US Twin Study of Age-Related Macular Degeneration. *Arch.Ophthalmol.* 2006;**124**:995-1001.

15. Francis PJ, Zhang H, Dewan A *et al*. Joint effects of polymorphisms in the HTRA1, LOC387715/ARMS2, and CFH genes on AMD in a Caucasian population. *Mol.Vis.* 2008;**14**:1395-400.

16. Seitsonen SP, Onkamo P, Peng G *et al*. Multifactor effects and evidence of potential interaction between complement factor H Y402H and LOC387715 A69S in age-related macular degeneration. *PLoS.ONE.* 2008;**3**:e3833.

17. Seddon JM, Reynolds R, Maller J *et al*. Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. *Invest Ophthalmol.Vis.Sci.* 2009;**50**:2044-53.

18. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. *Am.J.Ophthalmol.* 2001;**132**:668-81.

19. Kuo TY, Lau W, Collins AR. LDMAP: the construction of high-resolution linkage disequilibrium maps of the human genome. *Methods Mol.Biol.* 2007;**376**:47-57.

20. Purcell S, Neale B, Todd-Brown K *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am.J.Hum.Genet* 2007;**81**:559-75.

21. Barrett JC, Fry B, Maller J *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;**21**:263-5.

22. Meng X, Zhang C, Chen J *et al*. Cloning and identification of a novel cDNA coding thioredoxin-related transmembrane protein 2. *Biochem.Genet.* 2003;**41**:99-106.

23. Kralovicova J, Vorechovsky I. SERPING1 rs2511988 and age-related macular degeneration. *Lancet* 2009;**373**:461-2.

24. Spencer KL, Olson LM, Anderson BM *et al*. C3 R102G polymorphism increases risk of age-related macular degeneration. *Hum.Mol.Genet.* 2008;**17**:1821-4.

25. Jakobsdottir J, Gorin MB, Conley YP *et al*. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS.Genet.* 2009;**5**:e1000337.

Acknowledgements

This work was supported by the Macular Disease Society; the Macula Vision Research Foundation and the Brian Mercer Charitable Trust. We would like to thank Helen Griffiths and Katherine Beeson for technical assistance.

Legends to Figures;

Figure 1. The SERPING1 region analysed, showing the negative natural log of the p-values (-ln p-value) for each allelic chi square, the LDU map, haplotype blocks and the locations of all genes in the region.

This plot shows the region analysed around the SERPING1 gene. For the 61 genotypes available the black line with open circles shows the results of the allelic chi square tests with the negative natural log of the p-values plotted on the left y-axis. The red dashed line shows the threshold for significance after Bonferroni correction. The green line shows the LDU map plotted in linkage disequilibrium units on the right y-axis. Flat regions on the LDU map represent extensive LD and steps represent areas of recombination where LD and thus ancestral haplotypes tend to be broken up. Below the x-axis all genes in the region are plotted with arrows to show their direction of transcription, the SERPING1 gene is in red and the TIMM10 gene is in blue and both are in bold. At the top of the diagram the locations of the ten haplotype blocks are also plotted as blue bars.

Figure 2. Histograms of risk scores for cases, controls and for cases split by AREDS score.

The distributions of risk scores are plotted for controls (A), cases (B), and cases split by AREDs score, AREDS2 (C), AREDS3 (D), AREDS 4 (E). The total number in each class is given along with the mean and 95% confidence interval.

Table 1. Single SNP analysis of SERPING1 region, five of 61 single SNP results have p<0.05 after correction.

Gene	SNP	Allelic χ^2	Uncorrected p-value	Permutation p-value	Odds ratio	95% Confidence Interval
SERPING1 (Intron)	rs2511989	18.42	1.77E-05	1.10E-03	0.67	0.56-0.80
SERPING1 (promoter)	rs2509897	17.51	2.86E-05	1.60E-03	0.67	0.55-0.81
TNXDC16 (Intron)	rs11229109	11.89	5.65E-04	3.33E-02	0.73	0.60-0.87
SERPING1 (promoter)	rs2244169	11.58	6.66E-04	3.90E-02	1.38	1.14-1.65
SERPING1 (promoter)	rs2511990	11.36	7.51E-04	4.42E-02	1.37	1.14-1.65

Table 2. Omnibus haplotype results for the 10 haplotype blocks.

Block	χ^2	Degrees of freedom	Uncorrected p-value
7	19.96	2	4.63E-05
6	19.54	5	1.52E-03
1	9.45	2	8.88E-03
9	13.55	5	1.87E-02
3	9.89	3	1.95E-02
8	18.87	9	2.63E-02
4	10.12	4	3.85E-02
5	7.11	3	6.85E-02
10	7.46	4	1.13E-01
2	7.20	5	2.06E-01

Omnibus haplotype results show the significance of a unique effect from any haplotype within the block. The dashed line shows the threshold for significant p-values < 0.05 after Bonferroni correction (10 blocks/ tests)

Table 3. Haplotype specific results for the significant blocks six and seven.

Block	Haplotype	Haplotype frequency	Case:Control frequencies	χ^2	Uncorrected p-value	Permutation p-value	Odds ratio	95% Confidence Interval
Block 7	GCTCGTTGGG	0.418	0.366 : 0.470	20.78	5.13E-06	2.00E-04	0.65	0.54-0.78
Block 6	GAAAGGG	0.336	0.371 : 0.301	10.20	1.40E-03	7.89E-02	1.37	1.13-1.66
Block 7	ATTCGTCAGG	0.309	0.343 : 0.275	10.12	1.50E-03	8.14E-02	1.38	1.13-1.68
Block 6	GGGTCAG	0.231	0.204 : 0.258	7.56	6.00E-03	2.80E-01	0.74	0.59-0.91
Block 6	GAATCAG	0.153	0.133 : 0.173	5.74	1.66E-02	5.95E-01	0.73	0.57-0.94

The dashed line shows the threshold for significant p-values <0.05 after correction by permutation.

Table 4. Logistic regression model.

Variable	B	S.E.	Wald	df	Sig.	Odds Ratios	95% Confidence Interval		PAR
							Lower	Upper	
Constant	-10.483	1.117	88.034	1	6.43E-21	0.000	-	-	-
Age	0.110	0.013	70.410	1	4.82E-17	1.117	1.088	1.146	-
HTRA1_rs11200638	1.031	0.159	41.771	1	1.03E-10	2.803	2.051	3.832	26.84
CFH_rs1061170	0.892	0.174	26.208	1	3.07E-7	2.439	1.734	3.431	35.10
CFH_rs1065489	0.630	0.234	7.263	1	0.007	1.877	1.187	2.968	11.32
CFH_rs2019727	-0.642	0.254	6.372	1	0.012	0.526	0.319	0.866	42.05*
SERPING1_rs2509897	-0.318	0.146	4.719	1	0.030	0.727	0.546	0.969	17.47*
Gender (female=1;male=0)	0.433	0.202	4.533	1	0.032	1.542	1.039	2.289	21.89
C3_rs2230199	0.364	0.171	4.550	1	0.033	1.440	1.030	2.012	7.63
Smoking (yes = 1; 0=no)	0.597	0.281	4.533	1	0.033	1.818	1.049	3.150	10.58

*Show protective effect, but PAR calculated for the “risk allele”.

Population attributable risk (PAR)= Pr(OR-1)/(1+Pr(OR-1)), where OR is the odds ratio and Pr the prevalence of the risk factor in the general population, estimated from controls. Combined PAR = 87.01%.

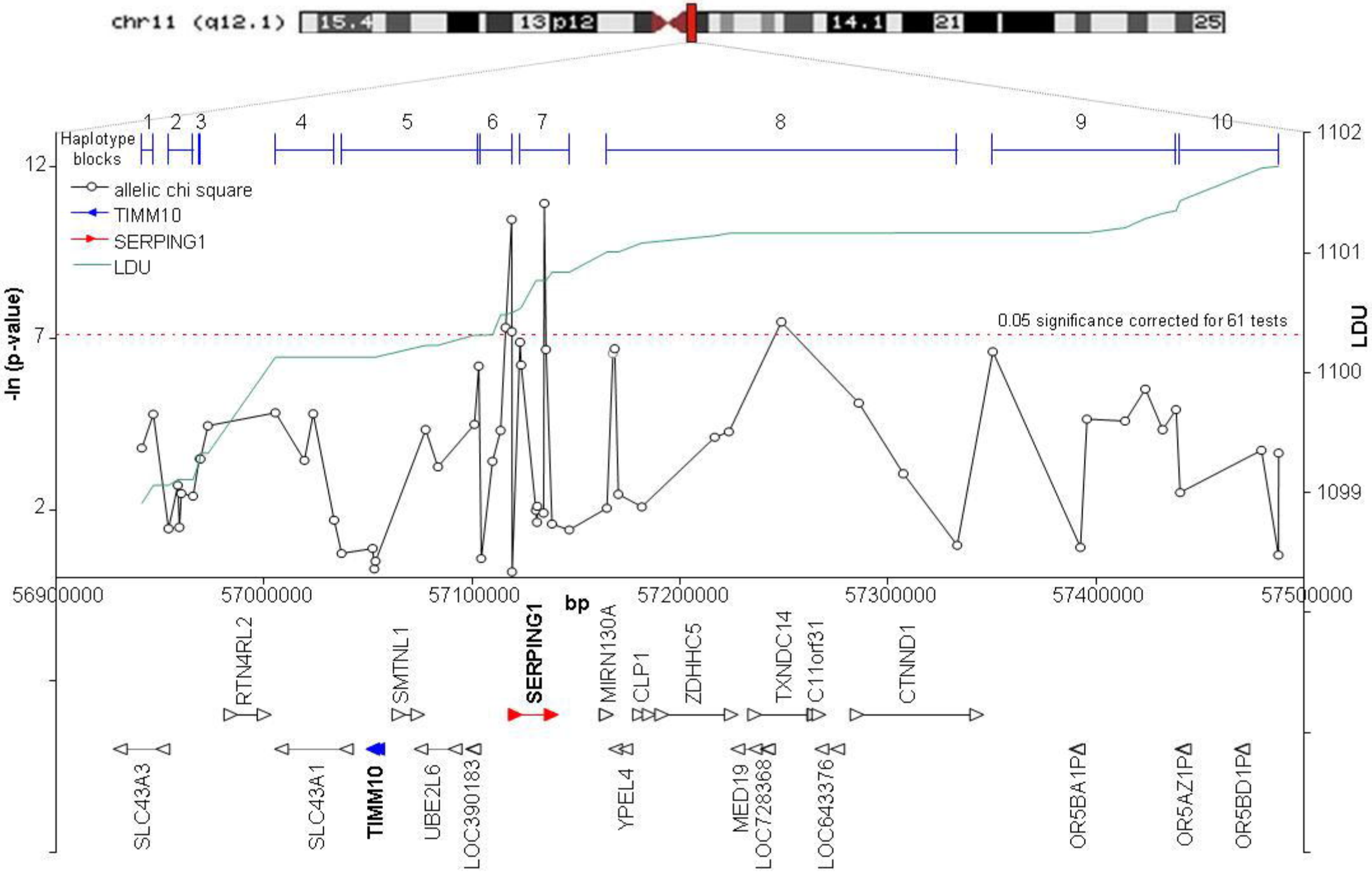
B = Logistic coefficient B.

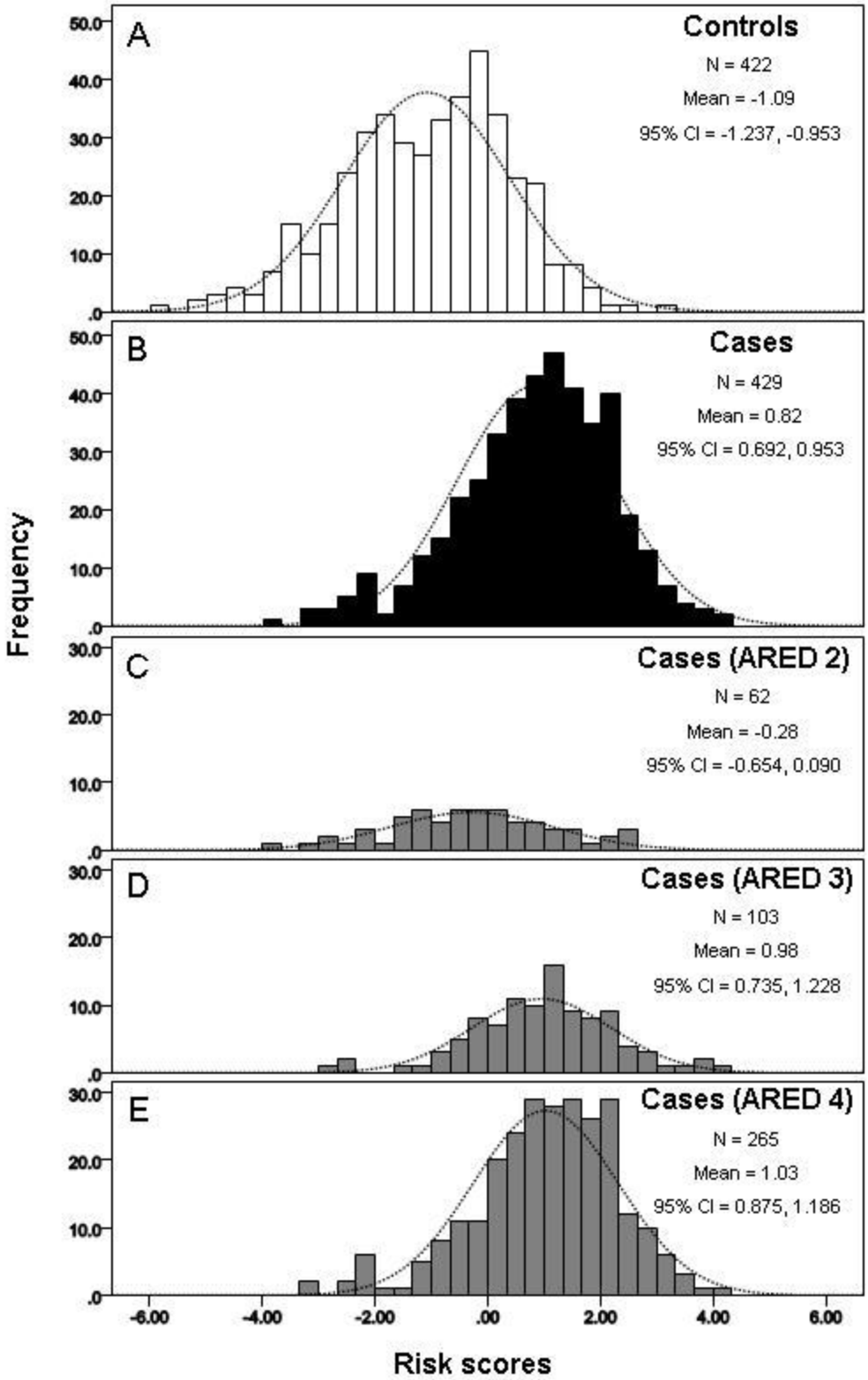
S.E. = Standard error of logistic coefficient B.

Wald = Wald chi square.

Sig. = p-value of Wald chi square.

df = Degrees of freedom.


chr11 (q12.1)
15.4
13 p12
14.1
21
25
Haplotype blocks
1
2
3
4
5
6
7
8
9
10
allelic chi square
TIMM10
SERPING1
LDU
0.05 significance corrected for 61 tests
-ln (p-value)
12
7
2
1102
1101
1100
1099
LDU
56900000
57000000
57100000
57200000
57300000
57400000
57500000
bp
SLC43A3
RTN4RL2
SLC43A1
TIMM10
SMTNL1
UBE2L6
LOC390183
SERPING1
MIRN130A
CLP1
YPEL4
ZDHHC5
MED19
LOC728368
TXNDC14
C11orf31
LOC643376
CTNND1
OR5BA1P
OR5AZ1P
OR5BD1P


A
Controls
N = 422
Mean = -1.09
95% CI = -1.237, -0.953
B
Cases
N = 429
Mean = 0.82
95% CI = 0.692, 0.953
C
Cases (ARED 2)
N = 62
Mean = -0.28
95% CI = -0.654, 0.090
D
Cases (ARED 3)
N = 103
Mean = 0.98
95% CI = 0.735, 1.228
E
Cases (ARED 4)
N = 265
Mean = 1.03
95% CI = 0.875, 1.186
Frequency
Risk scores
50.0
40.0
30.0
20.0
10.0
.0
-6.00
-4.00
-2.00
.00
2.00
4.00
6.00