

VARIANTS IN COMPLEMENT FACTOR PATHWAY GENES AND A NOVEL
LOCUS ON CHROMOSOME 16P12 INFLUENCE AGE-RELATED MACULAR
DEGENERATION SUSCEPTIBILITY

By

Kylee L. Spencer

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

December, 2007

Nashville, TN

Approved:

Professor Douglas P. Mortlock

Professor David E. McCauley

Professor Paul Sternberg

Professor Jeffrey A. Canter

Professor Jonathan L. Haines

To my parents and sister for always believing in me
Stephen R. and Darlene Littlepage
Stephanie Littlepage

and

To my husband for his unfailing support
William Clayton Spencer

ACKNOWLEDGEMENTS

I am greatly indebted to the members of my thesis committee (Dr Jonathan L. Haines, Dr. Jeffrey A. Canter, Dr. Douglas P. Mortlock, Dr. David McCauley, and Dr. Paul Sternberg). Thanks to Dr. Canter and Dr. McCauley for their statistical expertise, Dr. Mortlock for help working in a “pretty gnarly region” of the genome, and to Dr. Sternberg for sharing expert knowledge of the age-related macular degeneration phenotype.

I would especially like to thank my Ph.D. mentor, Dr. Jonathan L. Haines. Thanks for all of the guidance, time, and support. Most of all, thanks for teaching me how to think like a scientist.

I wish to thank the members of the Haines lab, past and present (Dr. Nathalie Schnetz-Boutaud, Brent Anderson, Melissa Allen, Becca Zuvich, Anna Davis, Dr. Jake McCauley, Dr. Sharon Liang, and Dr. Shannon Kenealy) for their contributions and dedication to all the ongoing projects in the lab and for friendships that made the journey enjoyable.

I have relied heavily on the expertise of everyone in the Center for Human Genetics Research, including the Computational Genomics Core (Chara Gibson and Lana M. Olson), Genetic Studies Ascertainment Core (Anthony Browning), and the DNA Resources Core (Cara Sutcliffe and Ping Mayo). Ms. Lana M. Olson is a first-rate statistician and was always willing to go above and beyond the call of duty to help me. Dr. Deborah Murdock was very generous in the use of her time and equipment, especially

for the CFHL3/CFHL1 deletion project. Thanks also to Dr. Chun Li for giving me a broader perspective on the field of statistics.

Of course, I owe many thanks to all my fellow graduate students for being there with me through this process. I would especially like to thank Todd Edwards and Digna Velez for serving as a sounding board whenever I needed to bounce around ideas and my friends in Cell and Developmental Biology for insight into genetics in model systems and all the good times we shared.

Finally, thanks to my parents, Stephen R. and Darlene Littlepage, and my sister Stephanie Littlepage, for all of their love and support over the years.

Most of all, thanks to Clay Spencer who is a great scientist, husband, and friend.

This work was supported by grants EY12118 and EY015216 from the NIH/National Eye Institute, grant M01 RR-00095 from the NIH/National Center for Research Resources, and the Genetics Training Program: Implications of Variation T32GM0652758. We are especially grateful to the study participants and clinicians without whom none of this work would have been possible.

TABLE OF CONTENTS

| | Page |
|---|------|
| DEDICATION | ii |
| ACKNOWLEDGEMENTS | iii |
| LIST OF TABLES | viii |
| LIST OF FIGURES | x |
| LIST OF ABBREVIATIONS..... | xiii |
| Chapter | |
| I. INTRODUCTION | 1 |
| Clinical Features of Age-related Macular Degeneration | 1 |
| Epidemiology | 4 |
| Genetic Epidemiology | 7 |
| Linkage Screens | 8 |
| CFH and the CFH-like Genes on Chromosome 1 | 9 |
| LOC387715 on Chromosome 10 | 17 |
| Candidate Gene Studies | 19 |
| CFB and CC2 on Chromosome 6 | 19 |
| C3 and APOE on Chromosome 19 | 20 |
| Conclusion | 21 |
| II. HYPOTHESES AND SPECIFIC AIMS | 22 |
| Specific Aim I | 24 |
| Specific Aim II | 24 |
| Specific Aim III | 25 |
| III. DELETION OF CFHL3 AND CFHL1 GENES IN AGE-RELATED MACULAR DEGENERATION | 26 |
| Introduction..... | 26 |
| Materials and Methods..... | 30 |
| Study Populations | 30 |
| SNP Genotyping | 31 |
| Deletion Screening by PCR and Agarose Gel Electrophoresis | 31 |
| Statistical Analysis..... | 32 |

| | |
|---|-----------|
| Results..... | 33 |
| Linkage Disequilibrium and Haplotype Association of the Pre-defined P1, P2, Risk, and Neutral Haplotypes | 33 |
| Segregation of the Deletion with CFH Haplotypes | 35 |
| Deletion Association Analyses | 38 |
| Discussion | 41 |
| | |
| IV. PROTECTIVE EFFECT OF COMPLEMENT FACTOR B AND COMPLEMENT COMPONENT 2 VARIANTS IN AGE-RELATED MACULAR DEGENERATION ... | 43 |
| Introduction..... | 43 |
| Materials and Methods..... | 45 |
| Study Populations | 45 |
| Genotyping..... | 48 |
| Statistical Analyses | 48 |
| Results..... | 50 |
| Discussion | 59 |
| | |
| V. NARROWING THE MINIMUM CANDIDATE REGION FOR THE AMD SUSCEPTIBILITY GENE ON CHR16P12 | 61 |
| Introduction..... | 61 |
| Materials and Methods..... | 64 |
| Ascertainment, Genotyping, and Quality Control | 64 |
| Linkage Disequilibrium | 65 |
| Statistical Analysis in the Family-based Dataset | 66 |
| Statistical Analysis in the Case-control Dataset | 67 |
| Analysis by Grade of AMD | 67 |
| Results..... | 68 |
| Linkage Disequilibrium | 68 |
| Linkage | 68 |
| Association..... | 74 |
| Discussion | 81 |
| | |
| VI. SELECTION AND TESTING CANDIDATE GENES ON CHROMOSOME 16P12 FOR LINKAGE AND ASSOCIATION WITH AMD | 83 |
| Introduction..... | 83 |
| Materials and Methods..... | 83 |
| Selection of Candidate Genes | 83 |
| Study Populations | 88 |
| SNP Selection and Genotyping..... | 93 |
| Statistical Analysis in the Family-based Dataset | 93 |
| Statistical Analysis in the Case-control Dataset | 95 |
| Results..... | 96 |
| Linkage Disequilibrium | 96 |

| | |
|--|-----|
| Linkage | 101 |
| Association Analysis for CACNG3 | 104 |
| Association Analysis for Q7Z6F8, HS3ST4, and IL4R | 105 |
| Discussion | 118 |
| | |
| VII. PREDICTIVE ACCURACY OF LOGISTIC REGRESSION AND IF-THEN DECISION TREE MODELS OF AMD | 121 |
| Introduction..... | 121 |
| Materials and Methods..... | 122 |
| Study Populations | 122 |
| Building the Models..... | 125 |
| Evaluating the Models | 127 |
| Results..... | 127 |
| Vanderbilt/Duke Study Population..... | 127 |
| Memphis Study Population..... | 134 |
| Discussion | 141 |
| Sample Size..... | 141 |
| Ascertainment Scheme..... | 142 |
| Missing Variables | 142 |
| | |
| VIII. CONCLUSIONS AND FUTURE DIRECTIONS..... | 144 |
| | |
| APPENDIX..... | 147 |
| | |
| REFERENCES | 155 |

LIST OF TABLES

| Table | Page |
|---|------|
| 1-1 Diagnostic AMD Grading | 3 |
| 3-1 Characteristics of the Deletion Screening Dataset | 31 |
| 3-2 Haplotype Association Analysis..... | 35 |
| 3-3 Univariate Association Analysis of the Deletion | 38 |
| 3-4 Features of Deletion Homozygotes | 39 |
| 3-5 Association Analysis of the Deletion in CFH Y402H TT Homozygotes..... | 40 |
| 3-6 Logistic Regression Analysis of the Deletion with Known AMD Risk Factors | 40 |
| 4-1 Characteristics of the Study Populations | 47 |
| 4-2 APL Association Results in the Family-based Dataset | 51 |
| 4-3 Allelic and Genotypic Association Results in the Case-control Dataset Comparing All AMD Cases (Grades 3,4, and 5) to All Controls (Grades 1 and 2)..... | 52 |
| 4-4 Allelic and Genotypic Association Results in the Case-control Dataset Comparing Neovascular AMD Cases (Grade 5) to Controls (Grade 1) | 53 |
| 4-5 Logistic Regression Analyses..... | 54 |
| 4-6 Conditional Analyses for CC2 E318D and CFB R32Q | 57 |
| 4-7 Comparison of Logistic Regression Models | 58 |
| 5-1 Characteristics of the Study Populations for the Screening Analyses..... | 64 |
| 5-2 Interesting Results form the Screening of the Broad Region of Interest on Chr16p12..... | 79 |
| 6-1 Characteristics of the 4 Candidate Genes | 87 |
| 6-2 Characteristics of the Study Populations | 89 |
| 6-3 Description of the SNPs Analyzed | 90 |

| | | |
|------|---|-----|
| 6-4 | Estimated Percentage of Alleles Captured Using SNPs Previously Genotyped Compared to Percentage of Alleles Captured by Adding 10 SNPs | 94 |
| 6-5 | Ordered Subset Analysis on Chromosome 16p12 | 103 |
| 7-1 | Comparison of the Two Halves of the Vanderbilt/Duke Study Population | 124 |
| 7-2 | Comparison of the Vanderbilt/Duke and Memphis Study Populations | 124 |
| 7-3 | Logistic Regression Model in Half the Vanderbilt/Duke Training Dataset | 128 |
| 7-4 | Evaluation of the Logistic Regression Model in the Vanderbilt/Duke Testing Dataset..... | 129 |
| 7-5 | If-Then Decision Tree Rules Developed in Half the Vanderbilt/Duke Training Dataset..... | 133 |
| 7-6 | Evaluation of the If-Then Decision Tree Rules in the Vanderbilt/Duke Testing Dataset..... | 134 |
| 7-7 | Logistic Regression Model in the Full Vanderbilt/Duke Training Dataset..... | 135 |
| 7-8 | Evaluation of the Logistic Regression Model in the Memphis Testing Dataset ... | 136 |
| 7-9 | If-Then Decision Tree Rules Developed in the Full Vanderbilt/Duke Training Dataset..... | 140 |
| 7-10 | Evaluation of the If-Then Decision Tree Rules in the Memphis Testing Dataset..... | 141 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1-1 Association of Risk Factors with Age-Related Macular Degeneration From Population-based Studies..... | 6 |
| 1-2 Odds Ratio Estimates for CFH Y402H | 10 |
| 1-3 Frequency of CFH Y402H in Various Ethnic Groups | 11 |
| 1-4 The Alternative Complement Cascade..... | 13 |
| 1-5 Frequency of the CFHL1/CFHL3 Deletion in Various Ethnic Groups..... | 15 |
| 1-6 Odds Ratio Estimates for LOC387715 A69S..... | 18 |
| 3-1 The Region of Complement Activation Gene Cluster..... | 28 |
| 3-2 Frequencies of Protective, Risk, and Neutral Haplotypes in Various Studies | 29 |
| 3-3 Genotyping the Deletion..... | 32 |
| 3-4 Linkage Disequilibrium in the Complete Dataset | 34 |
| 3-5 Haplotype Frequencies in the Deletion Homozygotes Compared to the Complete Dataset..... | 37 |
| 4-1 Classic, Lectin-binding, and Alternative Immune Pathways | 45 |
| 4-2 Linkage Disequilibrium for SNPs in CC2 and CFB on Chromosome 6 in the Case-control Dataset | 55 |
| 5-1 Preliminary Linkage and Ordered Subset Analysis on Chromosome 16..... | 63 |
| 5-2 Preliminary Association Analyses on Chromosome 16 | 63 |
| 5-3 Linkage Disequilibrium in the Family-based Dataset Chromosome 16p12 Grades 345vs12..... | 70 |
| 5-4 Linkage Disequilibrium in the Case-control Dataset Chromosome 16p12 Grades 345vs12..... | 71 |
| 5-5 Linkage Analysis in the Region of Interest on Chromosome 16p12..... | 72 |

| | | |
|------|--|-----|
| 5-6 | Ordered Subset Analysis on Chromosome 16p12..... | 73 |
| 5-7 | Association on Chromosome 16p12..... | 76 |
| 5-8 | Stratified Association Analysis Grades 345vs12..... | 78 |
| 6-1 | Selection of Candidate Genes..... | 85 |
| 6-2 | Linkage Disequilibrium in CACNG3 Case-control Grades 345vs12 | 97 |
| 6-3 | Linkage Disequilibrium in HS3ST4 Case-control Grades 345vs12..... | 98 |
| 6-4 | Linkage Disequilibrium in IL4R Case-control Grades 345vs12 | 99 |
| 6-5 | Linkage Disequilibrium in Q7Z6F8 Case-control Grades 345vs12..... | 100 |
| 6-6 | Linkage Analysis for the Candidate Genes on Chromosome 16p12..... | 102 |
| 6-7 | Family-based and Case-control Association Analysis in the Candidate Genes Grades 345vs12..... | 106 |
| 6-8 | Family-based and Case-control Association Analysis in the Candidate Genes Grades 5vs1 at Least One Eye | 108 |
| 6-9 | Family-based and Case-control Association Analysis in the Candidate Genes Grades 5vs1 Both Eyes | 110 |
| 6-10 | Stratified Case-control Association Analysis in the Candidate Genes Grades 345vs12..... | 112 |
| 6-11 | Family-based and Case-control Haplotype Analysis in the Candidate Genes Grades 345v12 | 114 |
| 6-12 | Duplication of PRKCB1 and CACNG3 | 119 |
| 7-1 | Two Approaches for Building the AMD Models..... | 125 |
| 7-2 | Histogram of Probabilities of AMD by Affection Status in the Duke/Vanderbilt Testing Dataset..... | 129 |
| 7-3 | ROC Curve in the Vanderbilt/Duke Testing Dataset | 130 |
| 7-4 | Counts of Cases and Controls for Each Susceptibility Combination in the Vanderbilt/Duke Training Dataset..... | 132 |
| 7-5 | Histogram of Probabilities of AMD by Affection Status in the Complete | |

| | | |
|-----|---|-----|
| | Vanderbilt/Duke Training Dataset..... | 137 |
| 7-6 | ROC Curve in the Memphis Testing Dataset..... | 138 |
| 7-7 | Counts of Cases and Controls for Each Susceptibility Combination in the Complete Vanderbilt/Duke Training Dataset | 139 |

LIST OF ABBREVIATIONS

| | |
|------------------|--|
| 2pt | two-point |
| ABI | Applied Biosystems |
| AMD | age-related macular degeneration |
| APL | association in the presence of linkage |
| APOE | apolipoprotein E |
| AREDS | Age-Related Eye Disease Study |
| ARM | age-related maculopathy |
| BDES | Beaver Dam Eye Study |
| BMI | body mass index |
| bp | base pair |
| CACNG3 | calcium channel, voltage-dependent, gamma subunit 3 |
| caco | case-control |
| CC2 | complement component 2 |
| CEPH | Centre d'Etude du Polymorphisme Humain |
| CFB | complement factor B |
| CFH | complement factor H |
| CFHL | CFH-like |
| CFHL1 | complement factor H-like 1 |
| CFHL3 | complement factor H-like 3 |
| chr | chromosome |
| CI | confidence interval |
| cM | centiMorgan |
| CNV | choroidal neovascularization |
| CRP | C reactive protein |
| DBP | diastolic blood pressure |
| DNA | deoxyribonucleic acid |
| Dom | dominant |
| DUMC | Duke University Medical Center |
| FARMS | Family ARM Study |
| GA | geographic atrophy |
| geno eff | genotyping efficiency |
| geno-PDT | genotype pedigree linkage disequilibrium test |
| Health ABC Study | Health Aging and Body Composition Study |
| Het-LOD | heterogeneity LOD score |
| HGDP | Human Genome Diversity Panel |
| HPFS | Health Professionals Follow-up Study |
| HS3ST4 | heparan sulfate (glucosamine) 3-O-sulfotransferase 4 |
| HTRA1 | HtrA serine peptidase 1 |
| HWE | Hardy-Weinberg Equilibrium |
| IL4R | interleukin 4 receptor |
| IOP | intraocular pressure |
| IVS | intervening sequence |
| LD | linkage disequilibrium |

| | |
|--------|---|
| LOD | logarithm of the odds |
| LRT | likelihood ratio test |
| MAF | minor allele frequency |
| Mb | megabase |
| MDR | Multifactor dimensionality reduction |
| Mx | multiplex |
| NCBI | National Center for Biotechnology Information |
| NEI | National Eye Institute |
| NHS | Nurses Health Study |
| NPLall | nonparametric linkage "Sall" statistic |
| OR | odds ratio |
| OSA | ordered subset analysis |
| P1 | CFH protective haplotype 1 |
| P2 | CFH protective haplotype 2 |
| PAR | population attributable risk |
| PCR | polymerase chain reaction |
| PDT | pedigree disequilibrium test |
| PED | pigment epithelial detachment |
| PRKCB1 | protein kinase C, beta 1 |
| RCA | region of complement activation |
| Rec | recessive |
| ROC | receiver operating characteristic curve |
| RPE | retinal pigment epithelium |
| SAGE | Serial Analysis of Gene Expression |
| SBP | systolic blood pressure |
| SNP | single nucleotide polymorphism |
| UTR | untranslated region |
| VUMC | Vanderbilt University Medical Center |

CHAPTER I

INTRODUCTION

Clinical Features of Age-related Macular Degeneration

Age-related macular degeneration (AMD) is a late-onset disorder characterized by central field vision loss, which may eventually lead to legal blindness. Three clinical hallmarks of AMD are the presence of large soft drusen, geographic atrophy (GA), and/or choroidal neovascularization (CNV). Other features associated with the disease include: pigmentary abnormalities, pigment epithelial detachment (PED), and disciform scarring.

Early in the disease process, soft drusen greater than 63 μm in diameter form between Bruch's membrane and the retinal pigment epithelium (RPE). Drusen represent undigested cellular debris, and often contain lipids and proteins such as vitronectin, TIMP3, crystallin, ubiquitin, integrins, factor X, APOE, and components of the complement system (Hageman et al. 2001). They vary from eye to eye in terms of shape, size, number, color, and distinctness of borders. The precise biological mechanism by which drusen form remains unknown, though several theories exist. One theory proposes that drusen are formed following injury to the retinal pigment epithelium as material from damaged cells leaks into the space between Bruch's membrane and the RPE (Hageman et al. 2001). Monocytes migrate to the site of this injury and mature into dendritic cells. The dendritic cells amplify the inflammatory response, producing a state of chronic inflammation. RPE cell debris and proteins secreted in response to the presence of the dendritic cells accumulate in drusen. After maturation, the dendritic cells leave the druse,

which continues to grow and soften. Alternatively, the hemodynamic model suggests that drusen in patients with wet AMD and higher capillary pressure are created due to a decreased ability of the RPE to transport waste against a higher hydrostatic gradient (Friedman 1997).

Whatever the mechanism, presence of large, soft drusen is an early warning sign that more severe features of AMD may develop. In the Beaver Dam Eye Study, late AMD was more likely to develop by the 5 year follow-up exam in individuals who had soft drusen at the baseline exam (6.5% vs. 0.1%) (Klein et al. 1997). In the Blue Mountain Eye Study, individuals with drusen $>63 \mu\text{m}$ at the baseline exam had an 11-fold increase in odds of CNV or GA at follow-up (Wang et al. 2003). However, many patients with early AMD are unaware of any visual deficits (Berger, Fine, and Maguire 1999), underscoring the need for detailed eye exams for controls, as well as cases, in genetic studies.

Geographic atrophy occurs when RPE cells die and their associated photoreceptors atrophy. GA may begin as a focal atrophy formed due to the regression of drusen near the fovea, out of an area of pigmentary abnormality, or from an RPE detachment. As the disease progresses, a horseshoe-shaped pattern of atrophy commonly surrounds the fovea, through which the choroidal vasculature is more easily observed. Finally, the fovea itself may begin to atrophy. It may take several years for this progression to occur, and patients often slowly notice a decrease in visual acuity (Berger, Fine, and Maguire 1999).

Geographic atrophy has been termed the “dry” form of late AMD, and occurs less frequently than the “wet” neovascular form.

“Wet” or exudative AMD is characterized by the presence of choroidal neovascularization, the growth of new blood vessels from the choroid through Bruch’s membrane into the subretinal space. Frequent hemorrhaging leads to the development of a disciform scar and severe vision loss. Unlike dry AMD, severe vision loss may occur suddenly after hemorrhaging.

Much debate exists in the field over how to define each subtype of AMD. One commonly used set of diagnostic criteria was developed by the Age-Related Eye Disease Study (AREDS 1999). Fundus photographs of our study participants have been graded by trained clinicians according to a modified version of this protocol, which is presented in Table 1-1.

Table 1-1. Diagnostic AMD Grading. Figure from Spencer et al 2007 *in press*.

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

Individuals with grade 3 AMD are at much higher risk to progress to geographic atrophy or choroidal neovascularization (Klein et al. 1997). Patients with GA may or may

not progress to CNV. Data from the Beaver Dam Eye Study and the Rotterdam Study have shown that choroidal neovascularization (grade 5 AMD) is about twice as common as geographic atrophy (grade 4 AMD) (Vingerling et al. 1995; Klein, Klein, and Linton 1992). This phenotypic heterogeneity has led some to suggest that perhaps "AMD" is a catchall term that covers several distinct disorders which share clinical features in common (Yanoff 2004).

Epidemiology

Based on a meta-analysis of population-based data from seven studies, the Eye Diseases Prevalence Research Group estimated that approximately 7.3 million people in the United States have large drusen characteristic of early AMD and that 1.75 million individuals suffer from advanced forms of this disorder (Friedman et al. 2004). The number of cases with geographic atrophy or choroidal neovascularization is expected to nearly double by the year 2020 as the population ages (Friedman et al. 2004). A meta-analysis of data from 3 large population-based studies of Caucasians in the United States, The Netherlands, and Australia estimated an overall prevalence of AMD of 0.2% in the 55-64 age group, which increased to 13% in those over 85 (Smith et al. 2001). The prevalence of late AMD in those of African or Hispanic descent seems to be lower than for Caucasians, and not much is known about the prevalence of AMD in other ethnic groups (Klein et al. 2004). A bias towards an increased number of affected Caucasian women vs. Caucasian men has been reported by some groups, but was not seen by others (Seddon and Chen 2004). Confounding due to the longer average life expectancy of women has made this gender bias difficult to confirm.

Age is the number one risk factor for developing AMD, followed by family history, smoking, and obesity. The connection between smoking and AMD has been particularly well-documented with 13 of 17 studies finding evidence of the association (Thornton et al. 2005). Meta-analysis of four prospective cohort studies estimated the odds ratio for current vs. never smokers at 2.35 (95% confidence interval 1.30 to 4.27) (Thornton et al. 2005). Obesity, defined as body mass index greater than 30, has been associated with both increased odds of central geographic atrophy (OR=1.93, 95% confidence interval 1.25 to 2.65) and progression to advanced AMD (RR=2.35, 95% confidence interval 1.27 to 4.34) (Clemons et al. 2005; Seddon et al. 2003a). Links between hypertension, iris color, and early menopause have been suggested, but were not associated with AMD in a meta-analysis of the Rotterdam, Beaver Dam, and Blue Mountain studies (Smith et al. 2001). The AREDS study also found no association between hypertension, iris color, or sun exposure (Clemons et al. 2005). Links between diabetes, alcohol consumption, cardiovascular disease, cataract surgery, and light exposure and AMD have also been suggested, though the results of these studies are inconclusive (Figure 1-1) (Klein et al. 2004).

TABLE. The Association of Risk Factors With Age-Related Macular Degeneration From Population-Based Studies*

| Study | History of Smoking | Blood Pressure | History of Diabetes | History of Alcohol Consumption | Overweight/Obesity | Body Mass Index | History of Cardiovascular Disease | UV Light Exposure | Light Exposure | Eye Color | Hyperopic Error | History of Cataract Surgery |
|---|--------------------|----------------|---------------------|--------------------------------|--------------------|-----------------|-----------------------------------|-------------------|----------------|-------------|-----------------|-----------------------------|
| NHANES I | | + (A) | | - Wine (A) | | | + Stroke, + (A) | | | | + (A) | |
| NHANES III | | + (E) | - (E) | + Heavy (E) | | | + (E) | | | | | |
| Framingham Eye study | O (A) | + (A) | O (A) | | | | + (A) | | | | + (A) | |
| Chesapeake Bay Waterman | O (L) | | | | | | | O (L) | ? + (L) | O | | |
| Copenhagen Heart study | + (A) | O (A) | | | | | O (A) | | | O | | |
| Beaver Dam Eye study | | | | | | | | | | | | |
| Prevalence | O (E), + (L) | O (E, L) | O (E), + (L) | O (E), + Beer (L) | + (E), O (L) | + (E), O (L) | O (E, L) | O (E, L) | O (E), + (L) | O (E, L) | O (E), + (L) | + (E), O (L) |
| Incidence | + (E), O (L) | O (E), + (L) | O (E, L) | O (E), + Heavy (L) | O (E, L) | O (E, L) | O (E, L) | O (E, L) | ? (E), O (L) | O (E, L) | O (E, L) | O (E), + (L) |
| Blue Mountains Eye study | | | | | | | | | | | | |
| Prevalence | + (E, L) | O (E, L) | O (E, L) | O (E, L) | + (E), O (L) | O (E, L) | O (E, L) | | | O (E, L) | O (E, L) | O (E, L) |
| Incidence | O (E), + (L) | O (E, L) | O (E, L) | + Heavy (L) | O (E, L) | O (E, L) | O (E, L) | | | O (E, L) | O (E, L) | O (E), + (L) |
| Rotterdam Eye study | | | | | | | | | | | | |
| Prevalence | + (E, L) | + (E), O (L) | O (E, L) | O (E, L) | O (E, L) | O (E, L) | O (E, L) | | | O (E) | O (L) | O (E, L) |
| Incidence | + (E, L) | + (A) | O (E, L) | O Heavy (E, L) | O (E, L) | O (E, L) | O (E, L) | | | + Brown (L) | + (A) | |
| Oulu study | O (A) | O (A) | O (A) | | | + A (male) | | | O (A) | | | O (A) |
| Colorado San Luis Valley | O (E) | O (E) | O (E) | -Wine (A) | | O (E) | + (E) | | | | | |
| French Age-related Eye Disease study (POLA) | + (L) | O (L) | O (E, L) | | + (L) | | O (L) | | O (L) | | | |
| Victoria, Australia (VP Study) | + (E, L) | | | | | O (A) | | | | | O (E, L) | |

A = any age-related macular degeneration; E = early age-related macular degeneration; L = late age-related macular degeneration; NHANES = National Health and Nutrition Examination Survey; UV = ultraviolet; + = positive association; O = no association; - = inverse association.

*Unless otherwise specified, associations are cross-sectional.

Source: Modified from Klein R. Epidemiology. In: Berger JW, Fine SL, Maguire MG, editors. Age-Related Macular Degeneration. St. Louis: Mosby, 1999:31-55, with permission from Mosby.

Figure 1-1. Association of Risk Factors with Age-Related Macular Degeneration From Population-based Studies. Figure from Klein et al. 2004.

Genetic Epidemiology

AMD is a complex, multifactorial late-onset disease. A strong genetic component in the etiology of AMD has long been suspected based on evidence from familial aggregation studies, twin studies, and segregation analysis.

Familial aggregation of the AMD phenotype has been well-documented in Caucasian populations. The prevalence of AMD in first degree relatives of AMD cases is higher than the prevalence in relatives of controls (23.7% vs 11.6%, respectively), with an age- and sex-adjusted odds ratio of 2.4 (95% confidence interval 1.2 to 4.7) (Seddon, Ajani, and Mitchell 1997). The Rotterdam Study also demonstrated familial aggregation of AMD and further showed that early manifestations of AMD may develop at a younger age in relatives of patients vs. relatives of controls (Klaver et al. 1998b). They estimated that the population-attributable risk related to genetic factors was at least 23%.

Furthermore, the Beaver Dam Eye Study examined younger siblings of AMD patients 5 years after the baseline exam in the probands to see if younger siblings had developed similar signs of AMD (Klein et al. 2001a). They found that the younger siblings were significantly more likely to develop soft drusen and RPE depigmentation within 5 years, if the older sibling had these lesions at the baseline exam. Recently, AMD was shown to cluster in Japanese families as well (Yoshida et al. 2000).

Numerous twin studies have suggested a substantial genetic component in risk for developing AMD (Seddon et al. 2005; Grizzard, Arnett, and Haag 2003; Hammond et al. 2002; Gottfredsdottir et al. 1999; Meyers, Greene, and Gutman 1995; Klein, Mauldin, and Stoumbos 1994; Meyers 1994). In the largest twin study to date of 840 male twins born between 1917 and 1927, the overall heritability of AMD was estimated at 46%

(Seddon et al. 2005). Heritability for intermediate AMD and late AMD was found to be 67% and 71%, respectively. This agrees remarkably well with the heritability estimate of 45% for overall AMD determined by Hammond et al. in their study of 516 female twin pairs with a mean age of 62(Hammond et al. 2002).

Finally, using segregation analysis, evidence for the presence of a major gene effect was obtained, and this study recommended a linkage screen as the next logical step for unraveling the genetics of AMD(Heiba et al. 1994).

Linkage Screens

As the name suggests, a genome-wide linkage screen takes an unbiased approach for localizing disease genes by testing a panel of markers spaced across the genome for linkage to the disease in pedigrees segregating the phenotype. One of the first genome-wide linkage screens analyzed one large family that appeared to be segregating an autosomal dominant form of AMD and implicated chromosome 1q25-q31(Klein et al. 1998). This region has turned out to be the most consistent finding of all the results of several genome scans(Weeks et al. 2001; Seddon et al. 2003b; Majewski et al. 2003; Weeks et al. 2004; Iyengar et al. 2004; Abecasis et al. 2004). Besides chromosome 1q25-q31, regions of linkage reported in 2 or more studies include 9q31, 10q26, 12q23, 15q21, 16p12, 17q25, and 22q13. Meta-analysis of the six largest genome screens strongly implicated 10q26, and supported linkage at 1q, 2p, 3p, 16p13-q12.2, and 16q12.2-q23.1(Fisher et al. 2005). Only the minimum candidate regions on 1q25-q31 and 10q26 have been sufficiently narrowed to associate specific genes with AMD, these being the

complement factor H and CFH-like genes on chromosome 1 and LOC387715/HTRA1 on chromosome 10.

CFH and the CFH-like Genes on Chromosome 1

To date the complement factor H gene remains the most widely accepted susceptibility locus for AMD. Our group reported odds ratios of 2.45 (1.41-4.25) for an individual carrying one risk allele and 3.33 (1.79-6.20) for carriers of two risk alleles. These increased to 3.45 (1.72-6.92) and 5.57 (2.52-12.27) when only neovascular AMD was examined(Haines et al. 2005). These odds ratios are very similar to what was reported by other groups, and to combined estimates from the largest meta-analysis of this polymorphism (Figure 1-2) (Schaumberg et al. 2007).

The population attributable risk (PAR) is a measure of the burden of disease placed on the population by a particular risk factor. We reported a PAR of 0.43(Haines et al. 2005), slightly lower than the range of 0.46-0.61 reported by other groups(Edwards et al. 2005; Klein et al. 2005; Zarepari et al. 2005a).

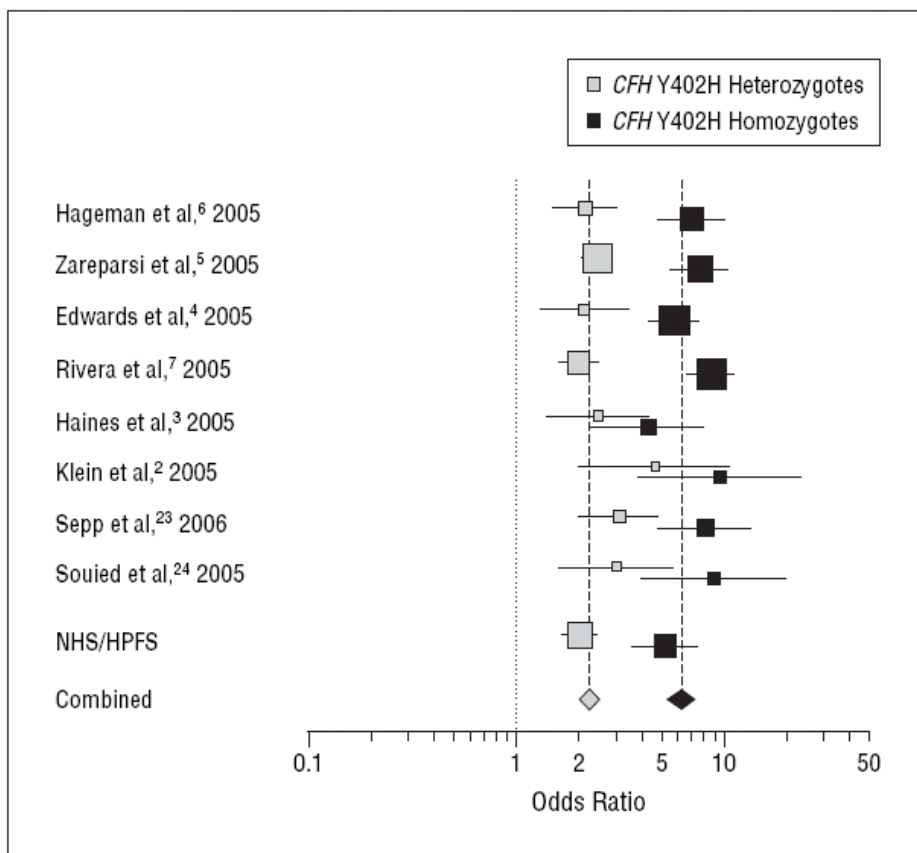


Figure 1-2. Odds Ratio Estimates for CFH Y402H. Figure from Schaumberg et al. 2007. NHS=Nurses Health Study, HPFS=Health Professionals Follow-up Study.

Interestingly, the minor allele frequency of CFH Y402H and the risk it carries for AMD vary across ethnic groups (Figure 1-3) (Hageman et al. 2006). Strong associations between CFH Y402H and AMD have been replicated in U.S. (Hageman et al. 2005; Conley et al. 2005; Zarepari et al. 2005a; Magnusson et al. 2006; Schaumberg et al. 2006; Seddon et al. 2006), U.K. (Sepp et al. 2006), French (Souied et al. 2005), German (Rivera et al. 2005), Icelandic (Magnusson et al. 2006), and Russian populations (Fisher et al. 2006). Haplotypes in CFH, but not the CFH Y402H polymorphism itself were associated with AMD in some Chinese study populations (Chen et al. 2006) and CFH Y402H was associated in others (Lau et al. 2006). This polymorphism was not associated

with early AMD in Latinos from Los Angeles, CA, though there was modest evidence of association with soft bilateral drusen (Tedeschi-Blok et al. 2007). Finally, no association with CFH Y402H was found in the Japanese, though risk and protective haplotypes within the gene were observed (Fuse et al. 2006; Gotoh et al. 2006; Okamoto et al. 2006; Tanimoto et al. 2006; Uka et al. 2006). Lack of statistical power could explain these negative results. However, the low overall frequency of CFH Y402H in the Japanese (~4%, Okamoto et al. 2006) coupled with consistent negative data across five studies, suggest that risk factors other than CFH Y402H are the major underlying contributors to AMD in this population.

| <i>CFH</i> T1277C | TT | CT | CC | f[C] |
|-------------------|-----|-----|----|-------|
| HGDP populations | | | | |
| European | 62 | 75 | 17 | 0.354 |
| African | 52 | 58 | 15 | 0.352 |
| Middle Eastern | 173 | 136 | 37 | 0.303 |
| North African | 15 | 14 | 1 | 0.267 |
| Pacific | 22 | 15 | 1 | 0.224 |
| Asian | 196 | 36 | 8 | 0.108 |
| Southeast Asian | 9 | 1 | 0 | 0.050 |
| North American | 90 | 5 | 0 | 0.026 |
| South American | 13 | 0 | 0 | 0.000 |

Figure 1-3. Frequency of CFH Y402H in Various Ethnic Groups. Figure adapted from Hageman et al. 2006. HGDP=Human Genome Diversity Panel.

The precise role of CFH in the AMD process is still being elucidated, but clues have been uncovered. It is known that CFH regulates the C3 component of the complement cascade, which left unchecked, may attack not only pathogens, but also host cells. CFH inhibits both initiation (by preventing assembly of active C3 convertase) and

amplification (by inactivating surface deposited C3b and destroying C3bBb convertase) of the complement cascade (Figure 1-4) (Zipfel et al. 2006). CFH also binds to anionic molecules on endothelial cells, marking them as self-tissues and preventing attack from the immune system (Alexander and Quigg 2007).

Using immunohistochemistry CFH protein has been observed in drusen, the subRPE space, and around the choroidal vasculature (Hageman et al. 2005). This staining is much more robust in AMD donor eyes than in age-matched controls. Mandal et al. confirmed localization of CFH in the retina and RPE-choroid, and further demonstrated that CFH expression in the eye increases with age in mice (Mandal and Ayyagari 2006).

There are many theories as to how CFH may be related to the pathophysiology of AMD. First, plasma levels of CFH are decreased in smokers (Esparza-Gordilla and Soria 2004), suggesting that smokers may be more susceptible to chronic inflammation of the retina and AMD. However, no statistical interaction between smoking and CFH Y402H genotype has been found (Scott et al. 2007; Seddon et al. 2006).

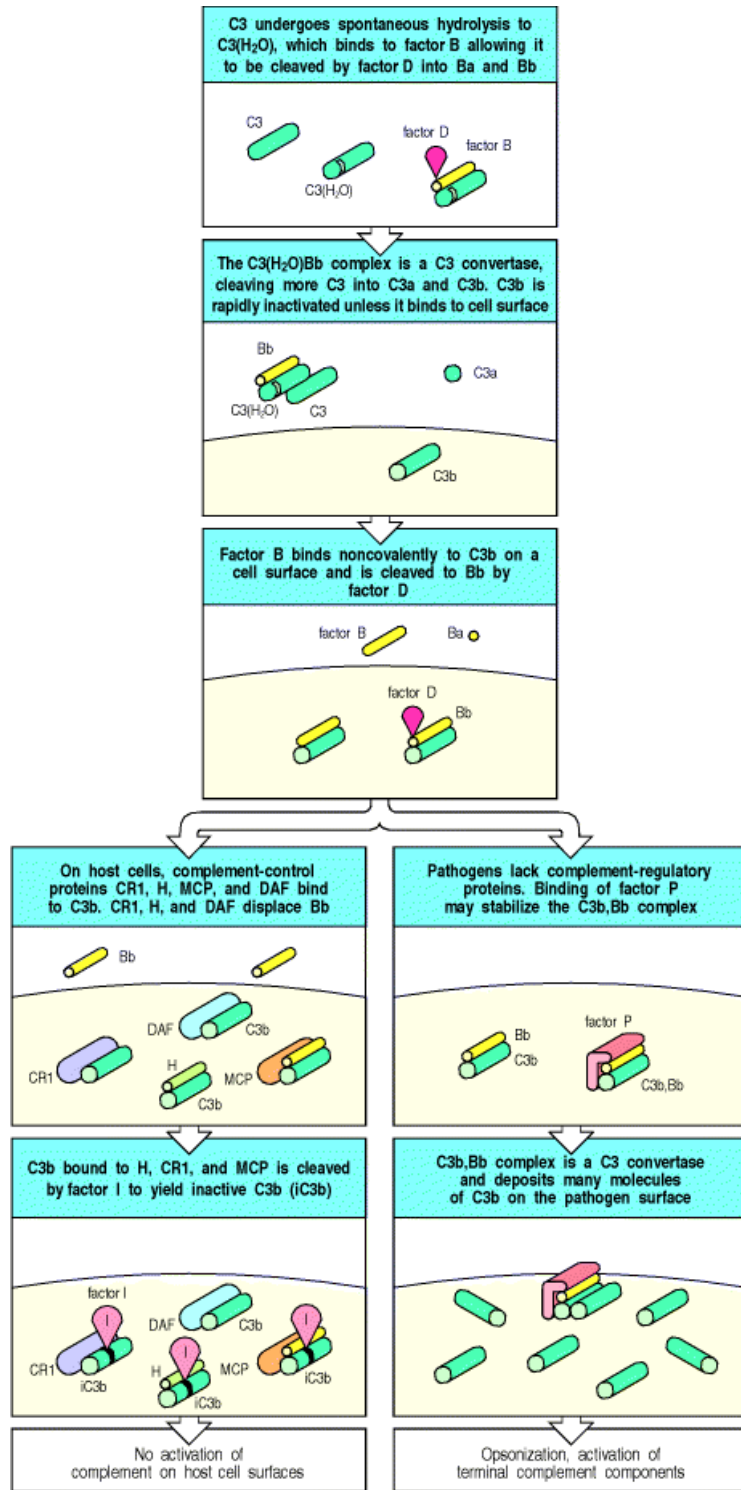


Figure 1-4. The Alternative Complement Cascade. CFH prevents activation of complement on host cells by inactivating surface deposited C3b (bottom left, top panel) and by acting as a cofactor for factor I so that C3b is cleaved to yield inactive C3b (bottom left, bottom panel). Figure from (Janeway, Jr. et al. 2001).

Second, the CFH Y402H polymorphism lies within the binding sites for C reactive protein (CRP) and heparin. When CFH binds to either CRP or heparin, its affinity for C3b is increased, and a stronger inhibition of the complement cascade results. Thus, this polymorphism may play a functional role by altering binding activity of the protein. Based on this evidence, it has been speculated that individuals with the high-risk CFH genotype have decreased ability to inhibit the complement system, and thus risk destruction of self-tissues. Recent functional data support decreased binding of CRP by the H402 version of the protein (Yu et al. 2007; Skerka et al. 2007; Laine et al. 2007). Additionally, histidine homozygotes have ~2.5 fold increase in CRP in the RPE-choroid compared to tyrosine homozygotes (Johnson et al. 2006).

Third, in addition to the risk imparted by the CFH Y402H allele, protective haplotypes spanning CFH exist (Hageman et al. 2005; Hageman et al. 2006)(Spencer et al. 2007, in press). An 84 Kb deletion of the CFH-like genes CFHL1 and CFHL3 segregates with one of these protective haplotypes (Hughes et al. 2006; Hageman et al. 2006), and CFHL1 and CFHL3 proteins were absent from the sera of deletion homozygotes. This led to the hypothesis that only the full-length transcript for CFH properly coordinates immune response and that the CFH-like proteins interfere with this regulation. Interestingly, the frequency of the deletion was highest in African Americans (Figure 1-5), possibly explaining the lower frequency of AMD in this population compared to Caucasians (7.4% of blacks vs. 15.8% of whites aged 75-84 years in the U.S.) (Hageman et al. 2006; Klein et al. 2006).

| <i>CFHR1/3</i> Deletion | +/+, +/-del | del/del | f[del/del] |
|-------------------------|-------------|---------|------------|
| US populations | | | |
| African American | 292 | 55 | 0.159 |
| Hispanic | 248 | 18 | 0.068 |
| European American | 266 | 13 | 0.047 |
| Chinese | 92 | 2 | 0.022 |
| HGDP populations | | | |
| African | 105 | 22 | 0.173 |
| North African | 24 | 5 | 0.172 |
| Middle Eastern | 180 | 31 | 0.147 |

Figure 1-5. Frequency of the CFHL1/CFHL3 Deletion in Various Ethnic Groups. Figure adapted from Hageman et al. 2006. HGDP=Human Genome Diversity Panel.

Besides AMD, polymorphisms in CFH are also associated with atypical hemolytic uremic syndrome (aHUS) and membranoproliferative glomerulonephritis type II (MPGN II). Symptoms of aHUS include hemolytic anemia, thrombocytopenia, and renal failure. CFH variants were found in ~24% of patients (Caprioli et al. 2003), and these individuals had a higher rate of kidney transplant failure and a higher mortality rate. Non-synonymous polymorphisms in the C-terminus of the CFH protein in aHUS patients have been associated with defective binding to anionic molecules on endothelial cells, leaving these cells open to attack from the immune system (Manuelian et al. 2003). Perhaps impaired binding of CFH to endothelial cells in the eye leads to the inflammation and tissue damage that is associated with AMD.

In MPGN II, dense deposits form in the glomerular capillary wall of the kidney, leading to defects in plasma filtration. Interestingly, many components of these dense deposits in the kidneys, such as vitronectin, immunoglobulin, and complement component C5, overlap with the constituents of drusen in AMD patients (Mullins et al. 2000). Furthermore, MPGN II patients may develop deposits in Bruch's membrane resembling

drusen in their teen years, much younger than the typical AMD patient (Duvall-Young, MacDonald, and McKechnie 1989).

Diseases like aHUS, MPGN II, and AMD provide an excellent illustration of phenotypic heterogeneity, in which more than one phenotype is derived from variation in a single gene. That all 3 diseases are associated with polymorphisms in CFH explains their similarities, but what accounts for the differences? Most obviously, the different diseases are related to different variations in CFH. For example, MPGNII is most often associated with mutations that cause decreased CFH plasma levels, whereas aHUS is usually related to mis-sense coding changes in the C-terminus of the protein, though there are exceptions (Alexander and Quigg 2007). Also, other genes are known to contribute independently of CFH to each of these phenotypes, and there may be other genetic modifiers that have not yet been discovered.

Finally, if CFH is such an important regulator of the complement system, then why don't changes in the gene cause systemic effects, rather than being concentrated mainly in the kidney and eye? Though no one knows for certain, it has been proposed that the great redundancy of complement regulatory proteins provides some protection when a single regulator is impaired and may limit defects to select tissues (Zipfel et al. 2006). If true, this redundancy could also help explain why it takes around 70 years for enough damage to accumulate in the macula to result in visual disturbances.

To summarize, CFH Y402H and other variation within CFH and the CFH-like genes has been strongly associated with AMD in a variety of populations. This discovery has provided insight not only into the pathophysiology underlying AMD, but also into related disorders, like aHUS and MPGN II. However, while CFH Y402H is highly correlated

with increased AMD risk, it is neither necessary nor sufficient to cause disease in all cases, hinting at the presence of other genetic and environmental susceptibility factors.

LOC387715 on Chromosome 10

Chromosome 10 showed the strongest evidence for linkage in a meta-analysis of the six largest genomic screens (Fisher et al. 2005), and the minimum candidate region was subsequently narrowed to just five genes: PLEKHA1, LOC387715, PRSS11 (aka HTRA1), GRK5, and RGS10 (Jakobsdottir et al. 2005). Several studies agree that the nonsynonymous LOC387715 A69S variant is associated with disease (Rivera et al. 2005; Schmidt et al. 2006; Schaumberg et al. 2007). The risk appears to translate across ethnic groups, with associations being reported not only in US, but also in Russian (Fisher et al. 2006) and Japanese populations (Tanimoto et al. 2006). In a meta-analysis of 3 studies, odds ratios were estimated at 2.18 for those carrying one risk allele and 7.14 in individuals homozygous for the risk allele (Figure 1-6). Cigarette smoking interacts with LOC387715 A69S genotype so that homozygous risk allele carriers who also smoke have an 8-22 fold increased risk of AMD (Schmidt et al. 2006; Schaumberg et al. 2007). Very little is known about the biological function of this gene, though expression has been observed in the retina (Rivera et al. 2005).

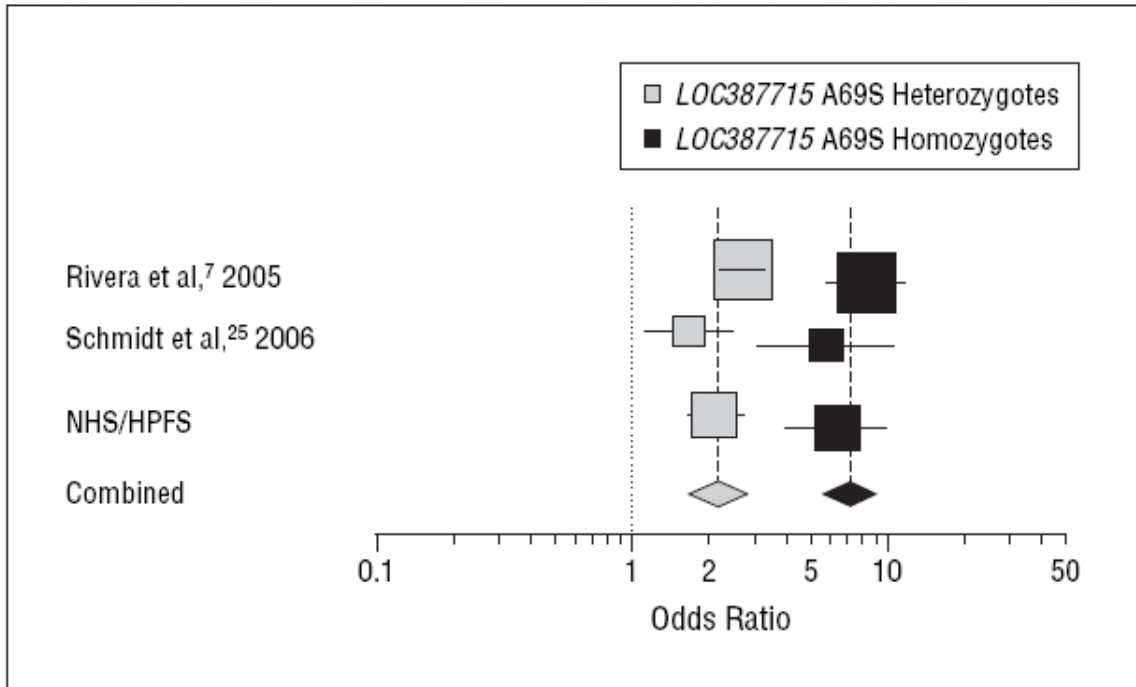


Figure 1-6. Odds Ratio Estimates for LOC387715 A69S. Figure from Schaumberg et al. 2007. NHS=Nurses Health Study, HPFS=Health Professionals Follow-up Study.

Conflicting reports suggest that LOC387715 A69S is not the functional variant in this region, but rather that it is a promoter SNP in HTRA1 that influences AMD susceptibility. This association was reported in Caucasian, Chinese, and Japanese populations (Dewan et al. 2006; Yang et al. 2006; Cameron et al. 2007; Yoshida et al. 2007; Mori et al. 2007). HTRA1 encodes a serine protease which regulates degradation of extracellular matrix proteoglycans. It has been conjectured that overexpression of HTRA1 may compromise Bruch's membrane, allowing blood vessels from the choroid to grow into the subretinal space (Yang et al. 2006). However, as LOC387715 A69S and the promoter SNP of HTRA1 are in very strong linkage disequilibrium, LOC387715 A69S can serve as a proxy for the risk variant in this region until the debate can be resolved.

Candidate Gene Studies

In contrast to genome scans, candidate gene studies involve scrutinization of a possible disease gene based on its known biology and test markers within the gene for association with disease. Association methods for both family-based and case-control data have been developed. Candidate gene studies to search for AMD susceptibility loci have focused mainly on genes that are involved in other macular disorders, extracellular matrix remodeling, oxidative stress, cardiovascular disease, cholesterol metabolism, or inflammation. To date these studies have remained largely unsuccessful, with initial association results either failing to be replicated at all or being replicated in some studies but not others. Of the genes identified using this type of approach, CFB, CC2, and APOE are the most consistently associated with AMD, and C3 is the most recently identified.

CFB and CC2 on Chromosome 6

After identification of CFH as a major AMD susceptibility locus, other genes in the complement pathways were explored. A strong association of the R32Q variant in CFB was observed in 3 independent Caucasian populations (Gold et al. 2006; Maller et al. 2006), and weaker association of the E318D variant in the adjacent CC2 gene was also reported (Gold et al. 2006).

CFB activates the alternative complement component pathway. CC2 is paralogous to CFB and activates the classical immune pathway. The CFB 32Q variant has reduced hemolytic activity compared to the CFB 32R variant, and it has been hypothesized that reduced CFB activity would prevent the chronic complement response that is thought to cause formation of drusen and progression to AMD (Gold et al. 2006). However, it is

unclear whether CC2 E318D, or another variant in linkage disequilibrium with this SNP, represents a second protective locus in this region.

C3 and APOE on Chromosome 19

Very recently, a non-synonymous polymorphism in complement component 3 (C3) was associated with increased risk of AMD in 2 populations from the United Kingdom (risk allele homozygotes OR=2.6, 95% confidence interval=1.6 to 4.1, PAR=22%), but to date replication of these results by independent groups has not been reported (Yates et al. 2007).

It has been hypothesized that altered lipoprotein metabolism may contribute to drusen formation (Anderson et al. 2001) and APOE has been found in drusen (Hageman et al. 2001). Five studies have shown that the APOE e4 allele has a protective effect (Zarepari et al. 2004; Van et al. 2004; Schmidt et al. 2002; Klaver et al. 1998a; Souied et al. 1998), and others have also seen suggestive evidence for a deleterious effect of the e2 allele (Zarepari et al. 2005b; Schmidt et al. 2002; Simonelli et al. 2001; Klaver et al. 1998a) or a lower age-at-onset in e2 carriers (Baird et al. 2004). Meta-analysis of four Caucasian datasets showed that APOE e4 was protective for both atrophic and neovascular AMD, regardless of gender (OR=0.54, 95% confidence interval=0.41 to 0.70). APOE e2 was only weakly associated with AMD, and only in men (OR for males =1.54, 95% confidence interval=0.97 to 2.45).

However, not all studies have seen a significant effect of APOE genotype with AMD (Nowak et al. 2004; Gotoh et al. 2004; Schultz et al. 2003; Schmidt et al. 2000; Pang et al. 2000). Possible reasons for this include: 1) lack of statistical power in small sample

sizes, 2) the APOE effect may be specific to a particular subgroup of AMD patients, which was not accounted for in these studies, or 3) this could be a true negative result. Because of the conflicting reports and the absence of association in our current dataset ($p > 0.14$ for both e2 and e4 alleles, after adjusting for age, smoking, CFH Y402H, LOC387715 A69S, and CFB R32Q), we have chosen not to include APOE as a covariate in our analyses.

Conclusion

To summarize, age, smoking, CFH Y402H, LOC387715 A69S, and CFB R32Q are confirmed AMD susceptibility factors. Though there are no published estimates of how many AMD cases can be attributed to these factors, variation in CFH, LOC387715, and cigarette smoking together explain approximately 61% of the population-attributable risk for AMD (Schmidt et al. 2006). While this is a significant portion, it implies that we still have not fully described how genetic and environmental variation contribute to this disease.

More work is needed to characterize the relationship between AMD and the CFHL1/CFHL3 deletion (explored in Chapter III) and to determine whether CC2 polymorphisms are independently associated with AMD (Chapter IV). To identify novel AMD loci, we should fine map linkage signals that replicate across studies, like chromosome 16p12 (Chapters V and VI). Finally, once data from all known AMD susceptibility factors is integrated, we can begin to see how well our hypothesized model of AMD applies to the general population (Chapter VII). Through these efforts, we hope to clarify the complex interplay of factors that contribute to AMD.

CHAPTER II

HYPOTHESES AND SPECIFIC AIMS

Age-related macular degeneration (AMD) is a complex, multifactorial late-onset disease that is the leading cause of blindness in the elderly. Large, soft drusen representing undigested cellular debris manifest early in the course of the disease. Later, central field vision loss occurs as geographic atrophy and/or choroidal neovascularization develop. Nongenetic or environmental risk factors suggested to play a role in AMD pathogenesis include increased age, smoking, and body mass index. A strong genetic component has long been suspected in the etiology of AMD based on the results of familial aggregation studies, segregation analyses, and numerous twin studies. Though several candidate gene studies have searched for a link between genes involved in oxidative stress, inflammation, and cardiovascular disease, associations between AMD susceptibility and variants in the complement factor H, LOC387715, complement factor B, and complement component 2 genes have been the most widely replicated.

Complement factor H inhibits activation of the alternative complement cascade, and thereby avoids injury to self tissues by preventing an excessive immune response. Inflammatory processes play a central role in AMD by contributing to the formation of drusen (Anderson et al. 2002), a clinical hallmark of AMD. Both risk and protective haplotypes in the CFH gene are known to modify susceptibility to AMD (Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005; Hageman et al. 2005). Along with CFH, five “CFH-like” genes reside within the regulator of complement activation (RCA) gene

cluster on chromosome 1. While the function of the CFH-like genes is largely unknown, the high degree of sequence similarity between these genes, and the suggestion that CFH and the CFH-like genes arose by duplication of a common ancestral sequence, point to an overlapping function of the CFH-like genes in innate immune system function and/or regulation. Hughes et al. proposed that deletion of the CFH-like genes CFHL1 and CFHL3 confers significant protection from AMD (Hughes et al. 2006). The putative deletion segregates completely with a particular haplotype of sequence variants in CFH. If replicated, this finding would yield substantial insight into how regulation of the complement system modulates susceptibility to AMD.

As variants in CFH are known to modify AMD susceptibility, it is only logical to ask whether polymorphisms in other genes within the alternative complement cascade also impart either risk or protection for AMD. CFB on chromosome 6 helps initiate activation of the alternative complement cascade. CC2 is paralogous to CFB and activates the classical component pathway. Variation in both CFB and CC2 has been associated with decreased risk for AMD (Gold et al. 2006), but it is unclear if the effect of CC2 E318D is independent of polymorphisms in CFB or caused by linkage disequilibrium between them (Maller et al. 2006). Dissection of this region will be essential for synthesizing a complete picture of AMD genetic susceptibility.

In addition to investigating how the complement system modulates AMD susceptibility, the search for novel AMD loci continues. Several genomic screens have been published, and a meta-analysis of the 6 largest genome screens has shown evidence of linkage on chromosome 16p12. In our own initial analysis a multipoint LOD score of 1.35 was obtained using a parametric dominant model in this region. Using Ordered

Subset Analysis (OSA) in our dataset, LOD scores between 2.0 and 2.9 were obtained after taking into account clinically meaningful covariates, such as systolic blood pressure, intraocular pressure (IOP), and body mass index (BMI). By using OSA to subset our families based on CFH Y402H LOD score, the LOD score on chromosome 16 increased to greater than 3. Association analysis in our family dataset and our independent case-control dataset also yielded encouraging results with several markers in this area producing p-values less than 0.01.

Therefore, my hypotheses are: 1) variation in the CFH-like, CFB, and CC2 genes is associated with AMD and 2) a susceptibility gene for AMD resides on chromosome 16p12. I propose the following specific aims:

Specific Aim 1: Investigate whether variation in the CFH-like genes, CFB, and CC2 contributes to AMD susceptibility. Replication of the putative deletion of CFHL1 and CFHL3 and the association of SNPs in CFB and CC2 in an independent dataset is essential. To that end, 3 tagging SNPs from the CFH haplotype will be genotyped in our full case-control dataset, and individuals homozygous for the CFHL1/CFHL3 deletion will be identified. PCR amplicons from CFH and CFHL1 visualized by agarose gel electrophoresis will confirm that individuals with the CFH haplotype also have the deletion. Fisher's exact test and logistic regression will be used to study the relationship between this deletion and AMD. Furthermore, 2 SNPs in CC2 and 4 SNPs in CFB will be genotyped and tested for association in both the family-based (PDT and APL) and case-control datasets (χ^2 test and logistic regression).

Specific Aim 2: Narrow the minimum candidate region for the AMD susceptibility gene on chromosome 16 using both linkage and association analysis. Approximately

200 SNPs will be genotyped in our family and case-control datasets on chromosome 16p12. SNPs will be selected based on informativeness, available assays, even spacing, and potential function. In our families 2pt and multipoint, parametric and nonparametric LOD scores will be calculated using Merlin to test for linkage in this region. Association analyses will be carried out using standard single marker chi-square tests of association in our case-control dataset. The Pedigree Disequilibrium Test (PDT), geno-PDT, and the “association in the presence of linkage” test (APL) will be used to study association in our families. Haploview will be used to determine the underlying linkage disequilibrium (LD) patterns in our datasets, and Haplo.stats and 2-SNP sliding windows in APL will be used to investigate association of particular haplotypes with AMD risk.

Specific Aim 3: Test candidate genes in the minimum candidate region on chromosome 16 for association with AMD. At least 5 additional SNPs in each of the locational/functional candidate genes identified in specific aim 2 will be genotyped. Association and LD analyses will be carried out in a similar fashion. Logistic regression and multifactor dimensionality reduction may be used to model gene-gene and gene-environment interactions, if a strong candidate locus is identified.

To summarize, the main goals of this proposal are to confirm the association of the CFHL1/CFHL3 deletion and CFB/CC2 polymorphisms with AMD and to identify a novel AMD susceptibility locus on chromosome 16p12.

CHAPTER III

DELETION OF CFHL3 AND CFHL1 GENES IN AGE-RELATED MACULAR DEGENERATION

Introduction

The genetic etiology of AMD, the leading cause of blindness in the elderly, is beginning to be unraveled with the discovery of the associations with Y402H in CFH on chromosome 1 and LOC387715 A69S on chromosome 10 (Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005; Jakobsdottir et al. 2005; Rivera et al. 2005). Taking into account known environmental/lifestyle risk factors for AMD has aided the search for predisposing genetic variants, as was the case when cigarette smoking was shown to act synergistically with the LOC387715 variant to further increase susceptibility (Schmidt et al. 2006). Furthermore, polymorphisms that reduce the risk of AMD are now being identified, including variants in factor B and complement component 2 on chromosome 6 (Gold et al. 2006; Spencer et al. 2007).

The CFH gene resides within the region of complement activation (RCA) gene cluster, which also includes 5 “CFH-like” or “CFH-related” genes (Figure 3-1). While the function of the CFHL genes is largely unknown, the high degree of sequence similarity between these genes and the suggestion that they arose out of duplication events with CFH point to an overlapping function of the CFHL genes in immune system function and/or regulation.

Haplotype analysis of CFH has revealed 2 common protective haplotypes and a neutral haplotype, in addition to the common risk haplotype carrying the C allele of CFH

Y402H (Figure 3-2) (Hageman et al. 2005; Hageman et al. 2006; Hughes et al. 2006; Spencer et al. 2007). However, synthesis of these results has been complicated since not every study used the same variants to define the haplotypes in the region. Nevertheless, from the frequencies of each haplotype and from the SNPs that do overlap between studies, it appears that P1, P2, and the neutral haplotype are the same across the 4 studies (Figures 3-2 and 3-5). Furthermore, if rs419137 is excluded from Hughes et al., then the risk 1 and risk 2 haplotypes merge into a single risk haplotype with a frequency of 61% in cases and 41% in controls, somewhat similar to the frequency of the risk haplotype in the other 2 Caucasian populations.

Interestingly, in the Japanese population the frequency of the risk haplotype is much lower than in Caucasians and is equal in cases and controls, suggesting that the C allele of CFH Y402H is not a major risk factor for AMD in the Japanese (Okamoto et al. 2006). This lack of association with CFH Y402H was later confirmed in other Japanese populations (Fuse et al. 2006; Gotoh et al. 2006; Tanimoto et al. 2006; Uka et al. 2006). Despite this, two haplotypes unique to the Japanese and not carrying the C allele at CFH Y402H, did significantly increase risk for AMD in this population, implying that while the risk associated with CFH Y402H does not translate across ethnic groups, variation in the CFH gene universally modulates susceptibility to AMD.

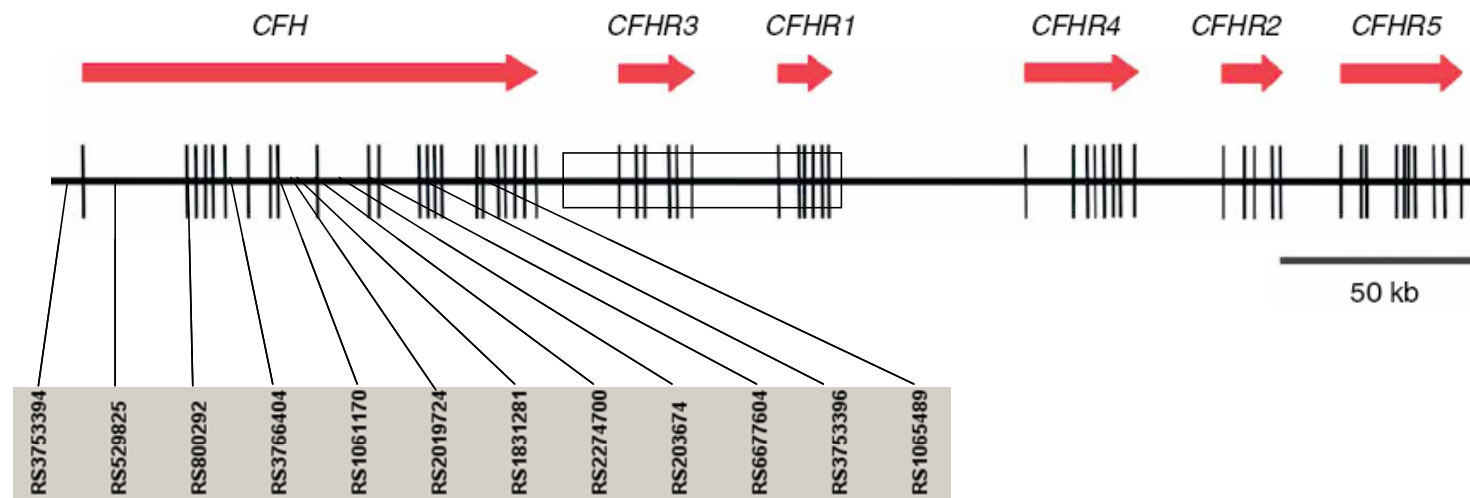


Figure 3-1. The Region of Complement Activation Gene Cluster. Exons are represented by solid vertical lines under each gene. The deletion of the *CFHL3* and *CFHL1* genes is indicated by the solid blue box. SNP positions and the deletion position are approximate. Adapted from Hughes et al. 2006.

| | rs3753394 | rs529825 | rs800292 | rs3766404 | rs1061170 | rs2019724 | rs203685 | rs1831281 | rs2274700 | rs203674 | rs6677604 | rs3753396 | rs419137 | rs2284664 | rs1065489 | Spencer et al. | | Hageman et al. | |
|---------|-----------|----------|----------|-----------|-----------|-----------|----------|-----------|-----------|----------|-----------|-----------|----------|-----------|-----------|----------------|----------|----------------|----------|
| | | | | | | | | | | | | | | | | Cases | Controls | Cases | Controls |
| P1 | C | A | A | T | T | | | | | T | | A | | | G | 0.13 | 0.20 | 0.12 | 0.21 |
| P2 | C | G | G | C | T | | | | | T | | A | | | G | 0.05 | 0.08 | 0.06 | 0.13 |
| Risk | C | G | G | T | C | | | | | G | | A | | | G | 0.51 | 0.38 | 0.50 | 0.29 |
| Neutral | T | G | G | T | T | | | | | T | | G | | | T | 0.14 | 0.14 | NR | NR |
| | | | | | | | | | | | | | | | | Okamoto et al. | | | |
| | | | | | | | | | | | | | | | | Cases | Controls | | |
| P1 | | | A | T | T | | | | | | | A | | | | 0.18 | 0.35 | | |
| P2 | | | G | C | T | | | | | | | A | | | | NR | NR | | |
| Risk | | | G | T | C | | | | | | | A | | | | 0.04 | 0.04 | | |
| Neutral | | | G | T | T | | | | | | | G | | | | 0.38 | 0.38 | | |
| | | | | | | | | | | | | | | | | Hughes et al. | | | |
| | | | | | | | | | | | | | | | | Cases | Controls | | |
| P1 | | | | | | C | T | A | A | | G | A | A | A | | 0.11 | 0.20 | | |
| P2 | | | | | | C | T | G | A | | A | A | A | G | | 0.08 | 0.19 | | |
| Risk 1 | | | | | | T | G | G | G | | G | A | A | G | | 0.43 | 0.29 | | |
| Risk 2 | | | | | | T | G | G | G | | G | A | C | G | | 0.18 | 0.13 | | |
| Neutral | | | | | | C | T | G | G | | G | G | A | G | | 0.20 | 0.18 | | |

Figure 3-2. Frequencies of Protective, Risk, and Neutral Haplotypes in Various Studies. Okamoto et al. studied a Japanese population; all others are Caucasian. We assumed based on the frequencies and the SNPs that overlap, that the P1, P2, Risk, and Neutral haplotypes are equivalent across the studies, but that is not certain. The alleles at rs2019724 and rs2274700 from Hughes et al. were switched to the opposite strand, so that they match the strand genotyped in this study. CFH Y402H=rs1061170.

It has been suggested that a deletion of ~84,000 basepairs covering the CFHL1 and CFHL3 genes which segregates with one of the protective CFH haplotypes is the source of the protective effect on chromosome 1 (Hughes et al. 2006; Hageman et al. 2006). We screened our dataset for deletion homozygotes and tested for association with decreased AMD susceptibility.

Materials and Methods

Study Populations

Our study population of 780 cases and 265 unrelated controls, all of Caucasian, non-Hispanic descent, was ascertained at Vanderbilt University Medical Center (VUMC) and Duke University Medical Center (DUMC) (Table 3-1). All patients and controls received an eye exam and had stereoscopic fundus photographs graded according to a modified version of the Age-Related Eye Disease Study (AREDS) grading system as described elsewhere (AREDS 1999; Schmidt et al. 2000). Grade 1 controls have no evidence of drusen or small non-extensive drusen without pigmentary abnormalities, while grade 2 controls may show signs of either extensive small drusen or non-extensive intermediate drusen and/or pigmentary abnormalities. Grade 3 AMD cases have extensive intermediate drusen or large, soft drusen with or without drusenoid retinal pigment epithelial detachment. Grade 4 AMD cases exhibit geographic atrophy and grade 5 individuals have exudative AMD, which includes nondrusenoid retinal pigment epithelial detachment, choroidal neovascularization, and subretinal hemorrhage or disciform scarring. Individuals were classified according to status in the more severely affected eye.

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

Table 3-1. Characteristics of the Deletion Screening Dataset

| Deletion Screening Dataset | | |
|-----------------------------------|----------------|--------------|
| | 780 | 265 |
| | cases | controls |
| | (grades 3,4,5) | (grades 1,2) |
| Total Individuals | | |
| Grade | 3: 25.9% | 1: 73.2% |
| | 4: 14.0% | 2: 26.8% |
| | 5: 60.1% | |
| Mean Age(sd) | 76.3(7.8) | 67.1(9.3) |
| % female | 64.2 | 57.7 |
| % ever smoked | 60.3 | 51.4 |

SNP Genotyping

Five SNPs used to define the deletion haplotype from Hughes et al. 2006 (rs2019724, rs1831281, rs2274700, rs6677604 and rs3753396) and 7 additional SNPs that make up the protective haplotypes identified by Hageman et al. 2005 (rs3753394, rs529825, rs800292, rs3766404, rs1061170, rs203674, and rs1065489) were genotyped according to the manufacturer's instructions using Taqman Assays on Demand from Applied Biosystems. Assays by Design were used when no pre-designed assay was available. Probe and primer sequences for the designed assays are available upon request.

Deletion Screening by PCR and Agarose Gel Electrophoresis.

We genotyped the deletion by PCR amplification with primers that amplify both a 325 bp product of CFH and a 381 bp product of CFHL1. Individuals with visible bands for

both amplicons by agarose gel electrophoresis were scored as not homozygous for the deletion. Individuals with the 325 bp band for CFH, but lacking the 381 bp band for CFHL1, were scored as being homozygous for the deletion (Figure 3-3).

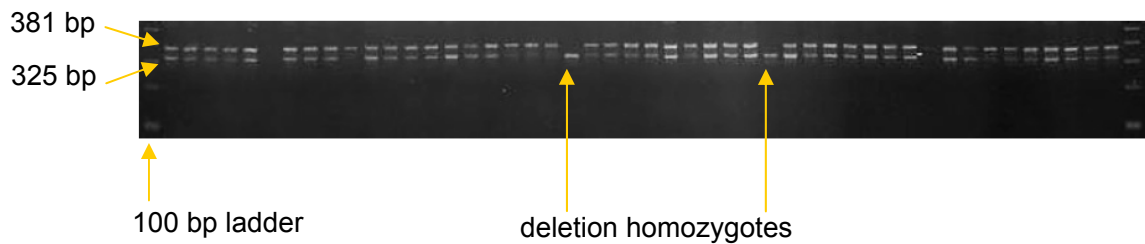


Figure 3-3. Genotyping the Deletion

The primer sequences were 5'-CTCTTCTTTTCTGCATCTGC-3' and 5'-ATTGCTGCTTATGGTAGATCAGG-3'. For each 10 ul reaction, 9.15 ul of Platinum PCR Supermix and 0.5 ul of each primer diluted with Puregene DNA Hydration Solution to a concentration of 0.1 ug/ul was added to 20 ng of genomic DNA. The PCR conditions were as follows: 1 cycle 96° C 15 min; 10 cycles 95° C 1 min, 55° C 1.5 min, 71° C 1 min; 30 cycles 95° C 1 min, 53° C 1.5 min, 71° C 1 min; 4° C hold. Agarose gel electrophoresis was used to visualize the resulting product(s) on a 2% agarose gel with ethidium bromide. A mix of all PCR reagents without a DNA sample added served as a negative control.

Statistical Analysis

Haploview software (Barrett et al. 2005) was used to examine linkage disequilibrium patterns in the complete dataset. Haplo.stats was used to estimate haplotype frequencies and generate score statistics for tests of association for the 12-SNP haplotype composed

of a blend of SNPs genotyped from the previous studies. Fisher's exact test was used to test for a significant difference in deletion homozygosity between all 780 cases and 265 controls and in a smaller group of 188 CFH Y402H TT homozygotes (Intercooled Stata 9.1 software, StataCorp LP, College Station, TX). Logistic regression was used to examine the effect of the deletion in the context of known AMD risk factors (SAS v9.1 software, SAS Institute, Cary, NC). The sample size for this analysis was reduced to 469 cases and 190 controls with complete age, CFH Y402H, LOC387715 A69S, deletion, and smoking data. Age was included in the model as a continuous covariate measured in years. CFH Y402H and LOC387715 A69S genotypes were coded as "1" for heterozygotes or homozygotes of the risk allele (CFH Y402H risk allele=C, LOC387715 A69S risk allele=T) and "0" for the non-risk allele homozygotes. Deletion homozygotes were coded as "1" and all others were coded as "0". Smokers (those who had smoked at least 100 cigarettes) were coded as "1" and non-smokers (those who had smoked fewer than 100 cigarettes over their lifetime) were coded as "0".

Results

Linkage Disequilibrium and Haplotype Association of the Pre-defined P1, P2, Risk, and Neutral Haplotypes

The 12 SNPs from the pre-defined haplotypes were in strong linkage disequilibrium with many r^2 values exceeding 0.20 (Figure 3-4). There are no strict rules for defining haplotype blocks, and though Haploview depicts two separate blocks of LD, we chose to include all 12 SNPs in one large haplotype across the region for ease of comparison with previous reports.

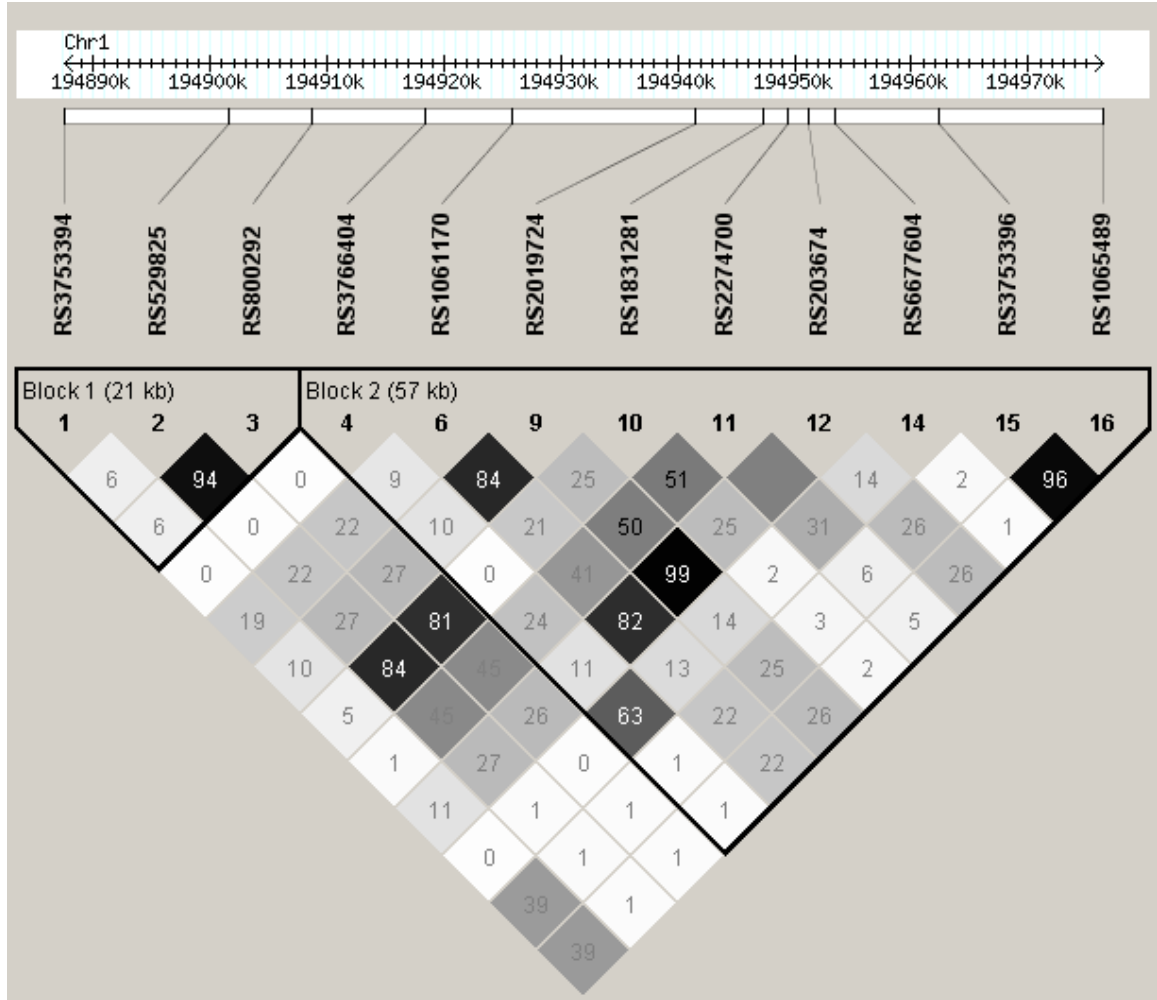


Figure 3-5. Linkage Disequilibrium in the Complete Dataset.

When testing the combined 12-SNP haplotypes for association, the P1 and risk haplotypes were very strongly associated with AMD, as expected from prior studies. ($p < 0.001$, Table 3-2). However, this set of SNPs did not capture the protective effect previously ascribed to P2 ($p = 0.10$, Table 3-2).

Table 3-2. Haplotype Association Analysis

| | Cases (%) | Controls (%) | Hap-Score | Empirical p-value |
|---------|------------------|---------------------|------------------|--------------------------|
| P1 | 0.12 | 0.21 | -4.58 | <0.001 |
| P2 | 0.05 | 0.07 | -1.62 | 0.10 |
| Neutral | 0.13 | 0.13 | -0.38 | 0.70 |
| Risk | 0.55 | 0.40 | 5.34 | <0.001 |

Segregation of the Deletion with CFH Haplotypes

After screening 1,045 samples, we identified 13 individuals who were homozygous for the deletion, 6 cases and 7 controls. Though the deletion segregated perfectly with the A allele of rs6677604 in previous studies (Hughes et al. 2006), this was not true in our dataset. While all of the deletion homozygotes were also homozygous for the A allele of rs6677604, three individuals who were not homozygous for the deletion were homozygous for the A allele of rs6677604.

The deletion seemed to segregate completely with alleles “GCGAAG” at SNPs rs529825, rs2019724, rs1831281, rs677604, rs3753396, and rs1065489 (red boxes, Figure 3-5), though with only 13 total deletion homozygotes observed, this could be an artifact of sample size. If we do use this haplotype as a marker for the deletion, then the overall estimated frequency of the deletion is ~10% (14% in controls, 8% in cases), and it follows Hardy-Weinberg Equilibrium ($p=0.42$). CFH protective haplotype 2 (P2) was the most frequent pre-defined haplotype found in these individuals with a frequency of ~47% (Figure 3-5), and the risk, neutral, and P1 haplotypes of CFH were not observed in this group. Co-segregation of the deletion with the risk allele of CFH Y402H was very rare (1 CT CFH Y402H heterozygote of 13 total deletion homozygotes), consistent with

previous reports (3 CT CFH Y402H heterozygotes and 3 CC homozygotes of 56 total deletion homozygotes, Hageman et al. 2006).

A.

| Deletion Homozygotes | | | | | | | | | | | | | |
|----------------------|----------|----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-------|----|
| RS3753394 | RS529825 | RS800292 | RS3766404 | RS1061170 | RS2019724 | RS1831281 | RS2274700 | RS203674 | RS6677604 | RS3753396 | RS1065489 | Total | |
| C | G | G | C | T | C | G | A | T | A | A | G | 0.47 | P2 |
| T | G | G | T | T | C | G | A | T | A | A | G | 0.29 | |
| T | G | G | C | T | C | G | A | T | A | A | G | 0.09 | |
| C | G | G | T | T | C | G | A | T | A | A | G | 0.07 | |
| T | G | G | T | C | C | G | G | G | A | A | G | 0.04 | |
| C | G | A | T | T | C | G | A | T | A | A | G | 0.02 | |
| C | G | A | C | T | C | G | A | T | A | A | G | 0.02 | |

B.

| Complete Dataset | | | | | | | | | | | | | |
|------------------|----------|----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-------|---------|
| RS3753394 | RS529825 | RS800292 | RS3766404 | RS1061170 | RS2019724 | RS1831281 | RS2274700 | RS203674 | RS6677604 | RS3753396 | RS1065489 | Total | |
| C | G | G | T | C | T | G | G | G | G | A | G | 0.51 | Risk |
| C | A | A | T | T | C | A | A | T | G | A | G | 0.14 | P1 |
| T | G | G | T | T | C | G | G | T | G | G | T | 0.13 | Neutral |
| C | G | G | C | T | C | G | A | T | A | A | G | 0.06 | P2 |
| T | G | G | T | C | T | G | G | G | G | A | G | 0.04 | |
| T | G | G | T | T | T | G | G | G | G | A | G | 0.03 | |
| T | G | G | T | T | C | G | A | T | A | A | G | 0.03 | |

Figure 3-5. Haplotype Frequencies in the Deletion Homozygotes Compared to the Complete Dataset. Alleles shared between all haplotypes and alleles shared between most haplotypes in the deletion homozygotes are marked in red and blue boxes, respectively. Notice, the overall frequency of P2 is 6%, but this increases to 47% in deletion homozygotes. The P1, risk, and neutral haplotypes are not found in the deletion homozygotes.

Deletion Association Analyses

Overall, deletion homozygosity was significantly more frequent in controls than cases (2.6% controls, 0.8% cases, Fisher's exact $p=0.025$, Table 3-3). Most of the deletion homozygotes were grade 1 controls (Table 3-4). Using the "GCGAAG" haplotype that all deletion homozygotes share as a surrogate for the deletion, this marker was strongly associated with AMD ($p<0.0001$, age-adjusted OR=0.40, 95% confidence interval=0.28 to 0.58).

Table 3-3. Univariate Association Analysis of the Deletion

a. 2x2 Contingency Table

| | AMD | Controls |
|----------------------|-----|----------|
| Deletion Homozygotes | 6 | 7 |
| All Others | 774 | 258 |

b. Association Results

| Fisher's exact p-value | Odds Ratio | 95% Confidence Interval | |
|---------------------------|---------------|-------------------------|------|
| 0.025 | 0.29 | 0.10 | 0.86 |

Table 3-4. Features of Deletion Homozygotes. ID=study participant identification number, M=male, F=female, Y=smoker, N=non-smoker, BMI=body mass index, “.”=missing data.

| ID | Gender | Affection Status | Grade | Age of Exam | Smoking | BMI | CFH Y402H | LOC387715 A69S |
|----|--------|------------------|-------|-------------|---------|------|-----------|----------------|
| 1 | F | Control | 1 | 58 | Y | . | TT | GT |
| 2 | M | Control | 1 | 68 | . | 25.1 | TT | . |
| 3 | F | Control | 1 | 61 | N | 44.1 | TT | GT |
| 4 | F | Control | 1 | 77 | Y | 24.4 | TT | GG |
| 5 | M | Control | 1 | 78 | N | 20.7 | TT | GG |
| 6 | F | Control | 1 | 61 | Y | 27.8 | TT | GT |
| 7 | F | Control | 2 | 64 | Y | 22.9 | TT | GT |
| 8 | F | Case | 3 | 69 | Y | 27.0 | TT | GG |
| 9 | M | Case | 3 | 78 | N | 25.7 | TT | GT |
| 10 | F | Case | 5 | 77 | N | 33.3 | CT | GG |
| 11 | F | Case | 5 | 90 | N | . | TT | GT |
| 12 | F | Case | 5 | 88 | . | 32.2 | TT | GG |
| 13 | F | Case | 5 | 71 | Y | 24.8 | TT | TT |

Because most of the deletion carriers are also homozygous for the non-risk allele of CFH Y402H, one may ask whether the deletion itself is truly protective or if the decreased risk is caused by absence of the CFH Y402H risk allele. To test this statistically, we stratified our dataset into CFH Y402H TT homozygotes and re-tested the deletion for association. Though the deletion was more than twice as frequent in controls compared to cases (9.9% controls, 4.3% cases), this difference was not statistically significant in the reduced sample of 188 CFH Y402H TT homozygotes (Fisher’s exact $p=0.22$, Table 3-5).

Table 3-5. Association Analysis of the Deletion in CFH Y402H TT Homozygotes**a. 2x2 Contingency Table**

| | AMD | Controls |
|----------------------|-----|----------|
| Deletion Homozygotes | 5 | 7 |
| All Others | 112 | 64 |

b. Association Results

| Fisher's exact p-value | Odds Ratio | 95% Confidence Interval | |
|---------------------------|---------------|-------------------------|------|
| 0.22 | 0.41 | 0.12 | 1.34 |

After controlling for age, CFH Y402H, LOC387715 A69S, and smoking, the protective effect of the deletion was no longer statistically significant (OR=0.45, 95% CI 0.11 to 1.83, p=0.27), though this is unsurprising given the low frequency of the deletion and the reduced sample size of 469 cases and 190 controls with complete covariate data (Table 3-6).

Table 3-6. Logistic Regression Analysis of the Deletion with Known AMD Risk Factors

| Effect | p-value | Odds Ratio | 95% Confidence Interval | |
|----------------|---------|------------|-------------------------|------|
| Age | <0.001 | 1.16 | 1.13 | 1.19 |
| Deletion | 0.27 | 0.45 | 0.11 | 1.83 |
| CFH Y402H | 0.03 | 1.72 | 1.07 | 2.79 |
| LOC387715 A69S | <0.001 | 2.83 | 1.88 | 4.26 |
| Smoking | 0.06 | 1.47 | 0.98 | 2.20 |

Discussion

Homozygosity for deletion of the CFHL3 and CFHL1 genes was inversely associated with AMD risk. Because of its low frequency, much larger sample sizes will be needed to test whether the deletion is protective in individuals without the CFH Y402H risk allele and in the context of other known AMD susceptibility factors. However, absence of the CFHL1 and CFHL3 proteins in the serum of deletion homozygotes and the hypothesis that CFHL1 and CFHL3 protein may compete with CFH for C3 binding and therefore interfere with normal regulation of the complement system by CFH argue against a purely statistical association (Hughes et al. 2006).

P2 was the most commonly observed pre-defined haplotype in deletion homozygotes. The frequency for P2 of 47% was slightly less than the 63% reported by Hageman et al., and much less than the 100% correspondence observed by Hughes et al., assuming that P2 is equivalent to their “Haplotype 5”. Without genotyping the other three SNPs used in that study, it is not possible to say the haplotypes are the same with certainty, but the alleles do match at the five SNPs that overlap between the two studies. Perhaps the perfect correspondence in the Hughes et al. study was caused by a founder effect in their Irish study population.

All deletion homozygotes were also homozygous for alleles “GCGAAG” at rs529825, rs2019724 rs1831281, rs6677604, rs3753396, and rs1065489 (red boxes, Figure 3-5). However, this combination of alleles should not be considered a complete surrogate for the deletion until a much larger screening of the population has confirmed that this is the case. Additionally, the risk, P1, and neutral haplotypes were not found in any of the deletion homozygotes.

Unfortunately, due to the nature of our genotyping assay, we were unable to distinguish between individuals with two copies of the CFHL1 and CFHL3 genes and deletion heterozygotes. We can overcome this limitation to some extent by using the “GCGAAG” haplotype as a marker for the deletion, though we do not know for certain that this haplotype marks the deletion in the population at large. Nevertheless, under this assumption, the deletion allele frequency can be estimated at ~14% in controls and 8% in cases, and it was strongly associated with decreased risk for AMD ($p < 0.0001$). In the future, we may be able to use more sophisticated methods, such as multiplex ligation-dependent probe amplification, to genotype deletion heterozygotes directly. It will be interesting to learn whether the deletion is completely tagged by the marker haplotype and also whether it is protective in the heterozygous state.

We have assumed that the breakpoints of the deletion we observed match those reported by Hughes et al. 2006, but this has yet to be confirmed in our dataset. Because of the high homology throughout this region, there are many opportunities for non-allelic homologous recombination to occur. In fact, a new fusion protein made from a CFH/CFHL1 hybrid gene has recently been discovered (exons 1-21 are derived from CFH and exons 22 and 23 from CFHL1) (Venables et al. 2006). Therefore, it is possible that other deletions, duplications, or hybrids may still be found.

In conclusion, deletion of CFHL1 and CFHL3 may account for a small portion of the protection from AMD associated with particular haplotypes in CFH. The presence of another protective haplotype that does not carry the deletion (P1, Table 3-2), suggests that other protective variants in this region have yet to be discovered.

CHAPTER IV*

PROTECTIVE EFFECT OF COMPLEMENT FACTOR B AND COMPLEMENT COMPONENT 2 VARIANTS IN AGE-RELATED MACULAR DEGENERATION

Introduction

Significant advances have been made in the field of age-related macular degeneration genetics with the identification of risk and protective haplotypes in the complement factor H (CFH) gene on chromosome 1 (Haines et al. 2005; Edwards et al. 2005; Klein et al. 2005) and the refined localization of the AMD risk gene on chromosome 10q26. Polymorphisms in both the poorly characterized LOC387715 gene and the serine protease HTRA1 have been strongly associated with AMD and replicated in multiple populations (Jakobsdottir et al. 2005; Rivera et al. 2005; Dewan et al. 2006; Yang et al. 2006; Yoshida et al. 2007). As these genes are adjacent on chromosome 10 and extensive linkage disequilibrium exists between them, identification of one or more “functional” variants will be difficult, but in the meantime the tagging SNP LOC387715 A69S in LOC387715 can be used as a proxy for the risk contributed by this genomic region. These associations continue to be refined as our knowledge of the interplay between genetic and lifestyle risk factors, such as cigarette smoking and increased body mass index, continues to grow (Schmidt et al. 2006; Schaumberg et al. 2007; Scott et al. 2007).

CFH inhibits activation of the alternative complement cascade, and thereby avoids injury to self tissues by preventing an excessive immune response. Inflammatory processes play a central role in AMD by contributing to the formation of drusen (Anderson et al. 2002), a hallmark feature of AMD in which deposits of extracellular

* This chapter adapted from Spencer et al. (2007) *Hum. Mol. Genet. In press.*

debris form between Bruch's membrane and the retinal pigment epithelium. Both risk and protective haplotypes in the CFH gene modify susceptibility to AMD (Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005; Hageman et al. 2005). Therefore, it is only logical to ask whether polymorphisms in other genes within the alternative complement pathway also impart either risk or protection for AMD.

Complement factor B (CFB) aids initiation of the alternative complement cascade, and complement component 2 (CC2) activates the classical component and lectin-binding pathways (Figure 4-1). CC2 is paralogous to CFB and resides adjacent to CFB on chromosome 6. Gold et al. demonstrated that variation in both CFB and CC2 is associated with decreased risk for AMD in two independent cohorts with a total of approximately 900 cases and 400 matched controls (Gold et al. 2006). Because of the strong linkage disequilibrium (LD) within the CFB/CC2 region, they were unable to completely determine which SNP(s) in these genes are the functional variant(s). Specifically, L9H in CFB, which is in strong LD with CC2 E318D, and CFB R32Q, which is in strong LD with an intronic SNP of CC2, were all highly protective for AMD in their study. Using stepwise logistic regression, Maller et al. (Maller et al. 2006) have excluded the intronic SNP of CC2 in favor of CFB R32Q. This agrees well with functional data showing that the CFB 32Q variant has reduced hemolytic activity compared to the CFB 32R variant and the hypothesis that reduced CFB activity would prevent the chronic complement response that is thought to cause formation of drusen and progression to AMD. However, ambiguity remains as to whether CFB L9H, CC2 E318D, or another variant in LD with these SNPs represents a second protective locus in this region. Further dissection of the CFB/CC2 region on chromosome 6 will be essential

for synthesizing a complete picture of the contribution of these loci to AMD genetic susceptibility.

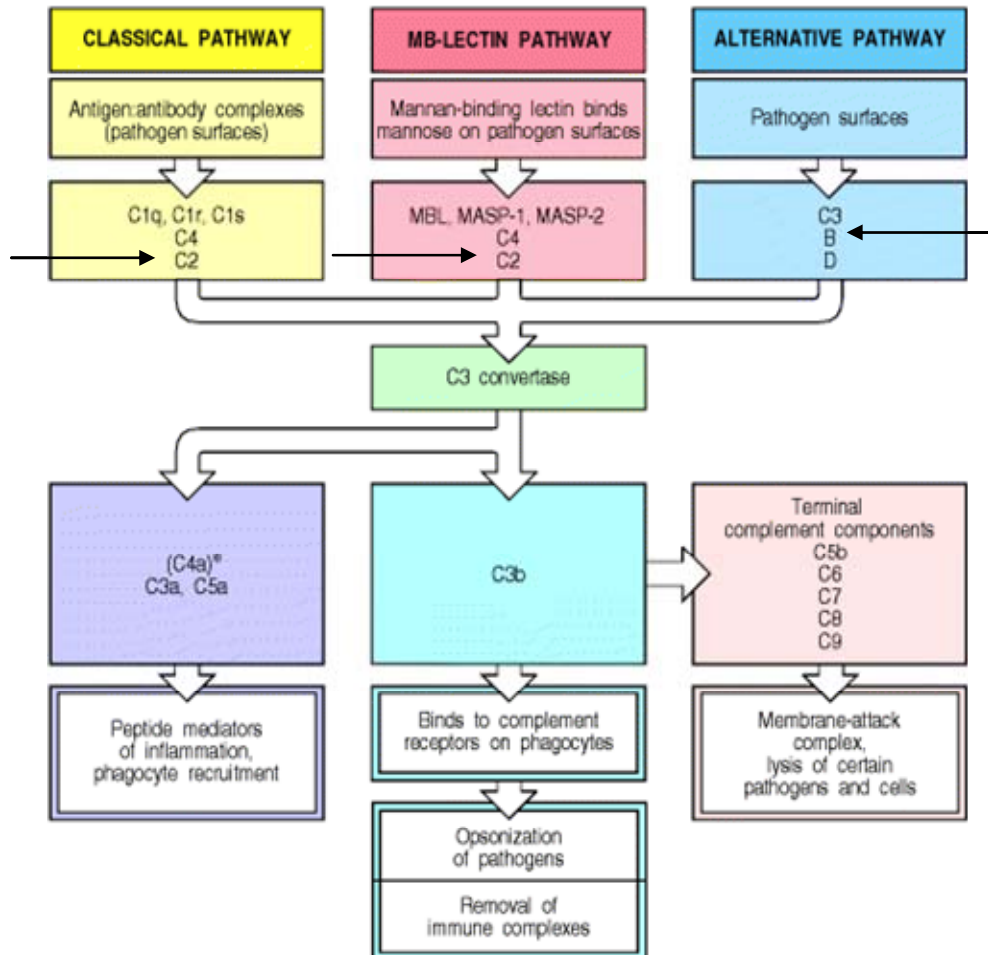


Figure 4-1. Classic, Lectin-binding, and Alternative Immune Pathways. Adapted from (Janeway, Jr. et al. 2001). Arrows highlight complement component 2 and complement factor B.

Materials and Methods

Study Populations

Multiplex and singleton families and an independent dataset of unrelated cases and controls, all of Caucasian, non-Hispanic descent, were ascertained at Vanderbilt

University Medical Center (VUMC) and Duke University Medical Center (DUMC). All patients and controls received an eye exam and had stereoscopic fundus photographs graded according to a modified version of the Age-Related Eye Disease Study (AREDS) grading system as described elsewhere (AREDS 1999; Schmidt et al. 2000). Briefly, grades 1 and 2 represent controls. Grade 1 controls have no evidence of drusen or small non-extensive drusen without pigmentary abnormalities, while grade 2 controls may show signs of either extensive small drusen or non-extensive intermediate drusen and/or pigmentary abnormalities. Grade 3 AMD cases have extensive intermediate drusen or large, soft drusen with or without drusenoid retinal pigment epithelial detachment. Grade 4 AMD cases exhibit geographic atrophy and grade 5 individuals have exudative AMD, which includes nondrusenoid retinal pigment epithelial detachment, choroidal neovascularization, and subretinal hemorrhage or disciform scarring. Individuals were classified according to status in the more severely affected eye. Table 4-1 describes additional features of the datasets, including age of exam, gender, and a brief description of family structure for the family-based dataset. Approval for the study was obtained from the appropriate institutional review boards at VUMC and DUMC, all study participants gave informed consent, and this research adhered to the tenets of the Declaration of Helsinki.

Table 4-1. Characteristics of the Study Populations. Mx=multiplex, SD=standard deviation.

| | Family Dataset | Independent Case-Control Dataset | |
|-------------------|--|----------------------------------|------------------------------|
| Total Individuals | 559 phenotyped (144 Mx, 79 Singleton families) | 698 cases (grades 3,4,5) | 282 controls (grades 1,2) |
| Grade | 1: 60.6% 3: 29.0% 2: 39.4% 4: 14.4% 5: 56.6% | 3: 27.4% 4: 12.6% 5: 60.0% | 1: 74.1% 2: 25.9% |
| Mean Age, SD | 67.3, 9.9 (unaffected) 74.7, 9.2 (affected) | 76.5 (7.6) | 66.7 (8.3) |
| % female | 66.5 | 63.6 | 55.3 |
| % ever smoked | 47.4 (unaffected) 56.2 (affected) | 60.9 | 49.6 |

Genotyping

Two SNPs in CC2 and four SNPs in CFB were selected for genotyping to validate the previously described association (Gold et al. 2006; Maller et al. 2006). SNPs rs9332739 (CC2 E318D), rs547154 (CC2 IVS10), rs1048709 (CFB R150R), and rs2072633 (CFB IVS17) were genotyped using Taqman Assays on Demand from Applied Biosystems (see Figure 4-2 for a diagram of this region). rs12614 (CFB R32W) and rs641153 (CFB R32Q) posed a more difficult problem for genotyping, as these SNPs are adjacent to each other on chromosome 6 (basepairs 32,022,158 and 32,022,159, respectively, NCBI Build 36). Taqman probes were designed to bind the 4 possible haplotypes for these two SNPs (CG, TG, CA, and TA). Each individual was then assayed with all possible combinations of the four probes (CA vic-labelled probe, TA fam-labelled probe; CG vic-labelled probe, CA fam-labelled probe; CG vic-labelled probe, TA fam-labelled probe; CG vic-labelled probe, TG fam-labelled probe; TG vic-labelled probe, CA fam-labelled probe; TG vic-labelled probe, TA fam-labelled probe), and the genotype determined by noting which combination of probes bound the DNA for each individual. Quality control samples were duplicated within and between plates, and we required that 95% of individuals assayed received a genotype for SNPs to be used in further analyses.

Statistical Analyses

We verified that all SNPs were in Hardy-Weinberg equilibrium (HWE) and examined the linkage disequilibrium between SNPs in both the family-based and case-control datasets using Haploview software (Barrett et al. 2005). HWE and LD in the case-control dataset were examined both in the overall dataset and separately in cases and controls.

The results were similar in each analysis (data not shown). We used only founders to estimate allele frequencies in the family-based dataset, except when a family did not have any founders genotyped. For those families, one individual was selected at random to contribute the allele frequency calculation. We tested SNPs for association in the family-based dataset using the “Association in the Presence of Linkage” (APL) method (Martin et al. 2003; Chung, Hauser, and Martin 2006). In the case-control dataset we assessed association of each SNP with AMD using a χ^2 test for allelic association and a 2x3 contingency table likelihood ratio test for genotypic association. We also used logistic regression in the case-control dataset to estimate the effects of CC2 E318D and CFB R32Q after controlling for age, smoking status, the CFH Y402H variant in complement factor H, and the LOC387715 A69S variant in LOC387715. Smokers (those who had smoked at least 100 cigarettes) were coded as “1” and non-smokers (those who had smoked fewer than 100 cigarettes over their lifetime) were coded as “0”. CFH Y402H and LOC387715 A69S genotypes were coded as “1” for heterozygotes or homozygotes of the risk allele (CFH Y402H risk allele=C, LOC387715 A69S risk allele=T) and “0” for the non-risk allele homozygotes. To conduct the conditional analyses, the case-control dataset was simply divided by minor allele carrier status for CC2 E318D or CFB R32Q, and the difference in allele frequency between cases and controls was tested by a χ^2 test, or Fisher’s exact test when the observed count in any cell was less than 5. We compared logistic regression models by calculating a likelihood ratio statistic (LRT, twice the difference in the deviance of the full compared to reduced logistic regression models) and determined significance by comparing the LRT to a χ^2 distribution with 1 degree of

freedom. All case-control analyses were performed using either SAS v9.1 software (SAS Institute, Cary, NC) or Intercooled Stata 9.1 (StataCorp LP, College Station, TX).

Results

To characterize the contribution of CC2 and/or CFB to AMD susceptibility we genotyped 6 SNPs in these genes in a family-based dataset with 559 individuals from 223 families, and a completely independent dataset of 698 cases and 282 unrelated controls (Table 4-1). All SNPs were in Hardy-Weinberg Equilibrium in the family dataset and in controls from the case-control dataset.

We observed weak evidence of association of the CFB R32Q SNP with decreased susceptibility to AMD in the family-based dataset, both in the overall analysis and when examining only neovascular AMD cases (APL $p=0.025$ overall, APL $p=0.014$ neovascular only, Table 4-2). In contrast, we saw strong association of SNPs in both CC2 and CFB in the case-control dataset in both the overall and neovascular AMD only analyses (Tables 4-3 and 4-4). As CC2 IVS10 and CFB R32Q are in strong linkage disequilibrium ($r^2=0.92$, Figure 4-2), we believe that these two SNPs are capturing the same protective effect. After controlling for age, CFH Y402H, and LOC387715 A69S, the minor alleles at both CC2 E318D and CFB R32Q were protective for AMD (Model 1, Table 4-5), though the association was much stronger for CFB R32Q than CC2 E318D (R32Q $p<0.0001$, odds ratio=0.29, 95% confidence interval 0.17 to 0.48; E318D $p=0.048$, odds ratio=0.48, 95% confidence interval 0.24 to 0.99).

Table 4-2. APL Association Results in the Family-based Dataset. MAF=minor allele frequency in founders. If a family did not contain any genotyped founders, one individual was selected at random to contribute to the allele frequency estimation.

| SNP | Gene | Location | bp | MAF | All AMD Grades 3,4, & 5 vs. Grades 1&2 p-value | Neovascular AMD Grade 5 vs. Grade 1 p-value |
|-----------|------|----------|----------|------|--|---|
| RS9332739 | CC2 | E318D | 32011783 | 0.03 | 0.221 | 0.447 |
| RS547154 | CC2 | IVS 10 | 32018917 | 0.07 | 0.241 | 0.208 |
| RS12614 | CFB | R32W | 32022158 | 0.09 | 0.988 | 0.633 |
| RS641153 | CFB | R32Q | 32022159 | 0.07 | 0.025 | 0.014 |
| RS1048709 | CFB | R150R | 32022914 | 0.18 | 0.134 | 0.638 |
| RS2072633 | CFB | IVS 17 | 32027557 | 0.42 | 0.123 | 0.419 |

Table 4-3. Allelic and Genotypic Association Results in the Case-control Dataset Comparing All AMD Cases (Grades 3,4, and 5) to All Controls (Grades 1 and 2)

| SNP | Gene | Location | Minor Allele | Overall MAF | All AMD MAF | All Controls MAF | All AMD Allelic p-value | All AMD Genotypic p-value |
|-----------|------|----------|--------------|-------------|-------------|------------------|-------------------------|---------------------------|
| RS9332739 | CC2 | E318D | C | 0.03 | 0.03 | 0.05 | 0.022 | 0.037 |
| RS547154 | CC2 | IVS 10 | T | 0.07 | 0.05 | 0.11 | 9.2E-06 | 5.9E-05 |
| RS12614 | CFB | R32W | T | 0.09 | 0.09 | 0.09 | 0.746 | 0.730 |
| RS641153 | CFB | R32Q | A | 0.06 | 0.05 | 0.10 | 2.3E-05 | 4.6E-04 |
| RS1048709 | CFB | R150R | A | 0.21 | 0.21 | 0.20 | 0.755 | 0.813 |
| RS2072633 | CFB | IVS 17 | A | 0.45 | 0.45 | 0.46 | 0.592 | 0.554 |

Table 4-4. Allelic and Genotypic Association Results in the Case-control Dataset Comparing Neovascular AMD Cases (Grade 5) to Controls (Grade 1). MAF=minor allele frequency

| SNP | Gene | Location | Minor Allele | Neovascular AMD MAF | Grade 1 Controls MAF | Neovascular AMD Allelic p-value | Neovascular AMD Genotypic p-value |
|-----------|------|----------|--------------|---------------------|----------------------|---------------------------------|-----------------------------------|
| RS9332739 | CC2 | E318D | C | 0.03 | 0.062 | 0.007 | 0.008 |
| RS547154 | CC2 | IVS 10 | T | 0.048 | 0.116 | 1.2E-05 | 1.7E-04 |
| RS12614 | CFB | R32W | T | 0.083 | 0.097 | 0.413 | 0.713 |
| RS641153 | CFB | R32Q | A | 0.044 | 0.114 | 3.6E-06 | 7.7E-05 |
| RS1048709 | CFB | R150R | A | 0.204 | 0.194 | 0.665 | 0.911 |
| RS2072633 | CFB | IVS 17 | A | 0.445 | 0.457 | 0.698 | 0.575 |

Table 4-5. Logistic Regression Analyses. The purpose of these analyses is two-fold: 1) to estimate the effect sizes of CC2 E318D and CFB R32Q in the context of accepted AMD risk factors, and 2) to provide a framework for likelihood ratio tests of model fit.

| Model | Cases | Controls | Effect | p-value | OR | 95% CI | |
|-------|-------|----------|---------------------------|---------|------|--------|------|
| 1 | 612 | 242 | age | <0.0001 | 1.17 | 1.14 | 1.19 |
| | | | rs1061170 CFH Y402H | 0.001 | 2.08 | 1.37 | 3.15 |
| | | | rs9332739 CC2 E318D | 0.048 | 0.48 | 0.24 | 0.99 |
| | | | rs641153 CFB R32Q | <0.0001 | 0.29 | 0.17 | 0.48 |
| | | | rs10490924 LOC387715 A69S | <0.0001 | 2.16 | 1.49 | 3.12 |
| 2 | 612 | 242 | age | <0.0001 | 1.17 | 1.14 | 1.19 |
| | | | rs1061170 CFH Y402H | 0.001 | 2.08 | 1.37 | 3.15 |
| | | | rs641153 CFB R32Q | <0.001 | 0.29 | 0.17 | 0.49 |
| | | | rs10490924 LOC387715 A69S | <0.001 | 2.21 | 1.53 | 3.19 |
| 3 | 612 | 242 | age | <0.001 | 1.16 | 1.13 | 1.19 |
| | | | rs1061170 CFH Y402H | <0.001 | 2.16 | 1.44 | 3.25 |
| | | | rs10490924 LOC387715 A69S | <0.001 | 2.23 | 1.55 | 3.19 |
| 4 | 400 | 204 | age | <0.0001 | 1.17 | 1.14 | 1.20 |
| | | | smoking | 0.057 | 1.51 | 0.99 | 2.32 |
| | | | rs1061170 CFH Y402H | 0.007 | 1.95 | 1.21 | 3.16 |
| | | | rs9332739 CC2 E318D | 0.263 | 0.60 | 0.25 | 1.47 |
| | | | rs641153 CFB R32Q | <0.0001 | 0.21 | 0.11 | 0.39 |
| | | | rs10490924 LOC387715 A69S | 3.0E-04 | 2.21 | 1.44 | 3.38 |
| 5 | 400 | 204 | age | <0.0001 | 1.17 | 1.14 | 1.20 |
| | | | smoking | 0.064 | 1.49 | 0.98 | 2.29 |
| | | | rs1061170 CFH Y402H | 0.006 | 1.96 | 1.21 | 3.17 |
| | | | rs641153 CFB R32Q | <0.0001 | 0.21 | 0.11 | 0.39 |
| | | | rs10490924 LOC387715 A69S | 2.0E-04 | 2.27 | 1.48 | 3.46 |
| 6 | 400 | 204 | age | <0.0001 | 1.16 | 1.13 | 1.19 |
| | | | smoking | 0.047 | 1.52 | 1.01 | 2.30 |
| | | | rs1061170 CFH Y402H | 0.002 | 2.07 | 1.30 | 3.30 |
| | | | rs10490924 LOC387715 A69S | <0.001 | 2.22 | 1.47 | 3.35 |

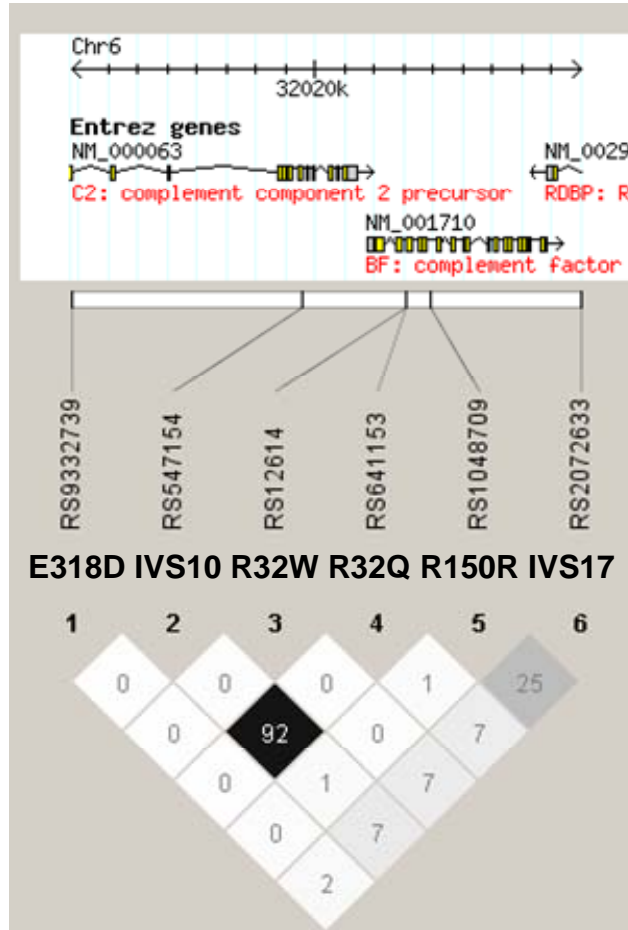


Figure 4-2. Linkage Disequilibrium for SNPs in CC2 and CFB on Chromosome 6 in the Case-control Dataset. Darker shading represents stronger linkage disequilibrium, and the r^2 values are shown inside each square. Results in the family-based dataset are quite similar (data not shown).

CFB R32Q and CC2 E318D do not exhibit strong linkage disequilibrium ($r^2=0.002$, Figure 4-2), and it's possible that both SNPs contribute independently to AMD susceptibility. Conditional analyses were used to further tease this apart. Conditioning on the C allele of CC2 E318D, the A allele of CFB R32Q is more frequent in controls than would be expected by chance, and vice versa (Table 4-6). For example, in carriers of the C allele of CC2 E318D, the frequency of the A allele of CFB R32Q is much higher in controls than cases (8% vs. 0%, $p=0.032$). In individuals that don't carry the C allele of CC2 E318D, the A allele of CFB R32Q is more frequent in controls than cases (11% vs. 5%, $p=2.3 \times 10^{-5}$). Therefore, CFB R32Q is associated with AMD regardless of genotype at CC2 E318D. In carriers of the A allele of CFB R32Q, the frequency of the C allele of CC2 E318D is higher in controls than cases (4% vs. 0%, $p=0.039$), and in individuals that don't carry the A allele of CFB R32Q, the C allele of CC2 E318D is again more frequent in controls than cases (5% vs. 3%, $p=0.034$). Therefore, CC2 E318D is associated with AMD regardless of genotype at CFB R32Q and vice versa, suggesting independent effects of both loci.

Table 4-6. Conditional Analyses for CC2 E318D and CFB R32Q. Freq=frequency, *p-value from Fisher’s exact test.

| | Freq of A allele of CFB R32Q in cases | Freq of A allele of CFB R32Q in controls | p-value |
|--|--|---|---------|
| Individuals carrying the C allele of CC2 E318D | 0.00 | 0.08 | 0.032* |
| Non-carriers of the C allele of CC2 E318D | 0.05 | 0.11 | 2.3E-05 |
| | Freq of C allele of CC2 E318D in cases | Freq of C allele of CC2 E318D in controls | p-value |
| Individuals carrying the A allele of CFB R32Q | 0.00 | 0.04 | 0.039* |
| Non-carriers of the A allele of CFB R32Q | 0.03 | 0.05 | 0.034 |

To further test the independent protective effects of CC2 E318D and CFB R32Q, we compared the fit of logistic regression models with and without these variables with a likelihood ratio test. Specifically, after adjusting for age, CFH Y402H, and LOC387715 A69S, a model including CFB R32Q fit the data significantly better than the model without CFB R32Q ($p < 0.001$, Model 2 vs. Model 3, Table 4-7). Furthermore, after adjusting for age, CFH Y402H, LOC387715 A69S, and CFB R32Q, a model including CC2 E318D fit the data significantly better than the model without CC2 E318D ($p = 0.049$, Model 1 vs. Model 2, Table 4-7).

Table 4-7. Comparison of Logistic Regression Models. LRT=likelihood ratio test, Models 1, 2, and 3 do not include smoking and have 612 cases and 242 controls. Models 4, 5, and 6 include smoking, and due to incomplete participant response to our lifestyle survey, the sample size is reduced to 400 cases and 204 controls.

| Comparison | Effect Removed from Reduced Model | LRT χ^2 | p-value |
|---------------------|-----------------------------------|--------------|---------|
| Model 1 vs. Model 2 | CC2 E318D | 3.88 | 0.049 |
| Model 2 vs. Model 3 | CFB R32Q | 21.5 | <0.001 |
| Model 4 vs. Model 5 | CC2 E318D | 1.25 | 0.26 |
| Model 5 vs. Model 6 | CFB R32Q | 25.2 | <0.001 |

Lastly, we examined the effect of these two SNPs in the context of smoking, in addition to controlling for age, CFH Y402H, and LOC387715 A69S. Due to incomplete participant response to the smoking questions on our lifestyle questionnaire, the sample size of this analysis was reduced from 612 cases and 242 controls with complete age and genotype data to 400 cases and 204 controls with complete data for age, genotypes, and smoking history. The effect of CFB R32Q remained strong in this analysis, but the evidence for association of CC2 E318D was much diminished (CFB R32Q $p < 0.0001$, odds ratio=0.21, 95% confidence interval 0.11 to 0.39; CC2 E318D $p = 0.26$, odds ratio=0.60, 95% confidence interval 0.25 to 1.47, Table 4-5 Model 4). Estimating the odds ratio for CFB R32Q controlling for age, CFH Y402H, LOC387715 A69S, and smoking without including CC2 E318D did not substantially alter the magnitude of this effect (CFB R32Q $p < 0.0001$, odds ratio=0.21, 95% confidence interval 0.11 to 0.39, Model 4, Table 4-5). When we compared models with and without CC2 E318D and CFB R32Q in our reduced sample with complete smoking data after adjusting for age, CFH Y402H, and LOC387715 A69S, the addition of CFB R32Q significantly improved model fit ($p < 0.001$, Model 5 vs. Model 6, Table 4-7), but the effect of CC2 E318D was no

longer statistically significant in the smaller sample ($p=0.26$, Model 4 vs. Model 5, Table 4-7).

Discussion

We have replicated the association of CFB R32Q and CC2 E318D in our case-control dataset. The conditional analyses and results of likelihood ratio testing suggest that these two polymorphisms exert independent, protective effects. However, after controlling for age, CFH Y402H, LOC387715 A69S, and smoking in logistic regression model 2, the effect of CC2 E318D was no longer statistically significant. This may be caused by reduced power to detect an effect in the smaller dataset with complete smoking covariate data. Reassuringly, CFB R32Q, which showed much stronger evidence for association in the allelic and genotypic association tests and in the conditional analysis, was still strongly associated with reduced AMD susceptibility after controlling for age, CFH Y402H, LOC387715 A69S, and smoking. The magnitude of this effect was not altered by including or excluding CC2 E318D variant from the model, which further suggests an independent effect of these two loci. However, statistical analyses can only shed so much light on genetic associations, and functional studies will be needed to unravel the mechanism behind these results.

To our knowledge, we are the first to study association of CFB and CC2 polymorphisms with AMD in a family-based dataset. We observed weak evidence for association of CFB R32Q in the families, but did not see an association of CC2 E318D. Lack of replication of CC2 E318D in the families could be a true negative result, but it is most likely due to reduced power, since this variant has a frequency of only 2.6% in our

families. One concern with any positive association result is the possibility that it derives from population substructure. One of the motivations for family-based studies is their insensitivity to population substructure, and our positive result for CFB R32Q in the family dataset provides further support that population stratification is not the cause of the observed association.

Finally, we have referred to the association of these variants with AMD as “protective” because the minor alleles of these polymorphisms are more frequent in controls than cases and to be consistent with previously published reports. However, since the actual biological mechanism by which these variants are acting to influence AMD susceptibility has not yet been described, and it is possible that the major allele may be acting as a risk allele, it may be more correct to term these results “inverse associations”, rather than “protective effects” until more is known about the underlying pathophysiology.

In summary, polymorphisms in CC2 and CFB are associated with protection from AMD, but future functional studies will be needed to confirm that these variants are the source of the decreased AMD susceptibility.

CHAPTER V

NARROWING THE MINIMUM CANDIDATE REGION FOR THE AMD SUSCEPTIBILITY GENE ON CHR16P12

Introduction

Much progress has been made in uncovering genetic modifiers of AMD, the leading cause of blindness in the elderly in developed countries. Even with the identification of CFH Y402H, LOC387715 A69S, CFB R32Q, CC2 E318D, and the CFHL1/CFHL3 deletion, other AMD susceptibility loci are hypothesized to exist. Fine mapping under strong linkage peaks from genome-wide screens is one strategy being applied to narrow this search.

Chromosome 16p12 was first proposed to harbor an AMD susceptibility locus based on the results of a genome-wide linkage screen in 263 sib pairs from 102 pedigrees in the Beaver Dam Eye Study (BDES) (Schick et al. 2003). Strong linkage was observed across a 6 cM region centered at ~18cM (26 Mb) (D16S679 $p=0.0086$).

Interestingly, in this analysis AMD was treated as a quantitative trait by using the 15-level Revised Wisconsin Age-Related Maculopathy Coding Protocol (Schick et al. 2003). This system considers: 1) drusen size, type, and area, 2) pigmentary abnormalities, 3) geographic atrophy, and 4) signs of exudative AMD to assign a severity score to each individual. While the scales are not directly comparable, roughly speaking, levels 4-11 correspond to grade 3 AMD in our study, level 12 is most similar to our grade 4, and levels 13-15 are roughly equivalent to our grade 5. The obvious drawback of this system is the cost of achieving high inter-rater reliability (usually by having pictures graded at a

specialized reading center), but this is offset by a gain in power from not “falsely dichotomizing” a trait with an inherent range of severity.

Using this same quantitative trait definition of AMD, linkage to D16S679 was replicated in the Family ARM Study (FARMS) of 349 sib pairs from 34 extended families ($p=0.0046$) (Iyengar et al. 2004). In contrast to the BDES, a community-based study with the full range of AMD severity scores represented, FARMS probands were ascertained from a retinal clinic and typically had severe AMD. Since linkage to the same marker appeared in two independent populations using different ascertainment schemes, we have increased confidence that chromosome 16p12 harbors a true susceptibility locus.

Furthermore, meta-analysis of the six largest AMD screens confirmed this locus ($p=0.0195$ for the 30 cM bin ranging from 17-52 Mb) (Fisher et al. 2005). Interestingly, there was some evidence of heterogeneity between studies at this locus (heterogeneity $p=0.012$), suggesting that only a subset of families may be linked to this region.

The heterogeneity hypothesis agrees very well with our own initial analysis, in which a multipoint LOD score less than 1.0 rose to between 2.0-3.0, after using ordered subset analysis to consider important clinical covariates (Figure 5-1) (Schmidt et al. 2004). Linked families on average tended to have higher systolic blood pressure, higher intraocular pressure (IOP), higher body mass index (BMI), and a lower proportion of affected individuals carrying the CFH Y402H risk allele. Association analyses in both the family-based and case-control datasets were also encouraging (several markers achieving nominal significance, p -values less than 0.05, Figure 5-2). There were 2 main clusters with interesting results in both the case-control and family datasets, centered at ~24 Mb and ~28-30 Mb.

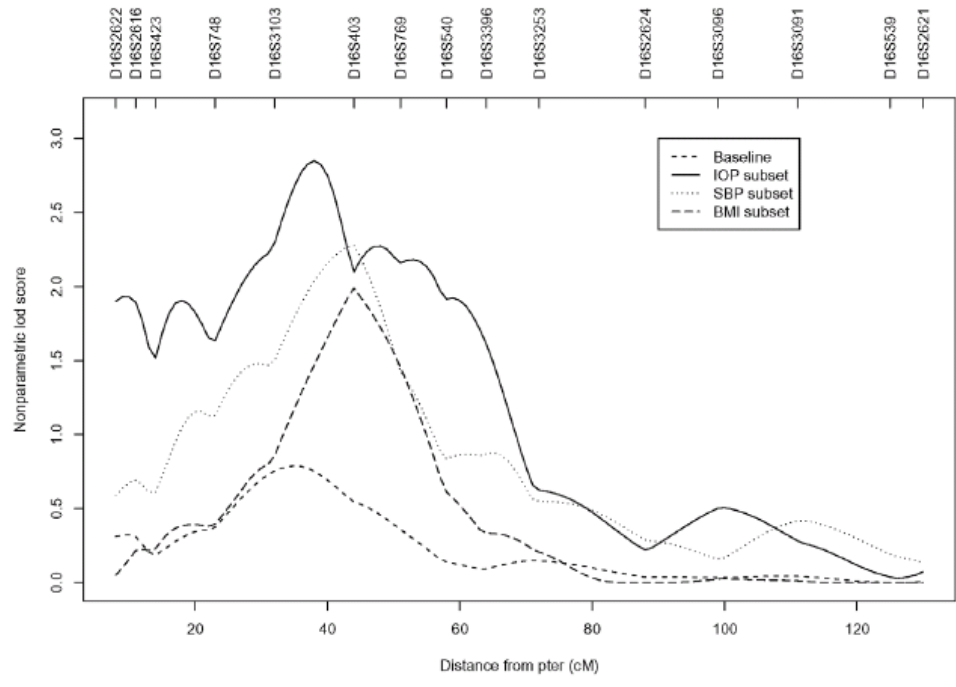


Figure 5-1. Preliminary Linkage and Ordered Subset Analysis on Chromosome 16. Figure from Schmidt et al. 2004. Build 36 D16S748=12Mb, D16S3253=53Mb.

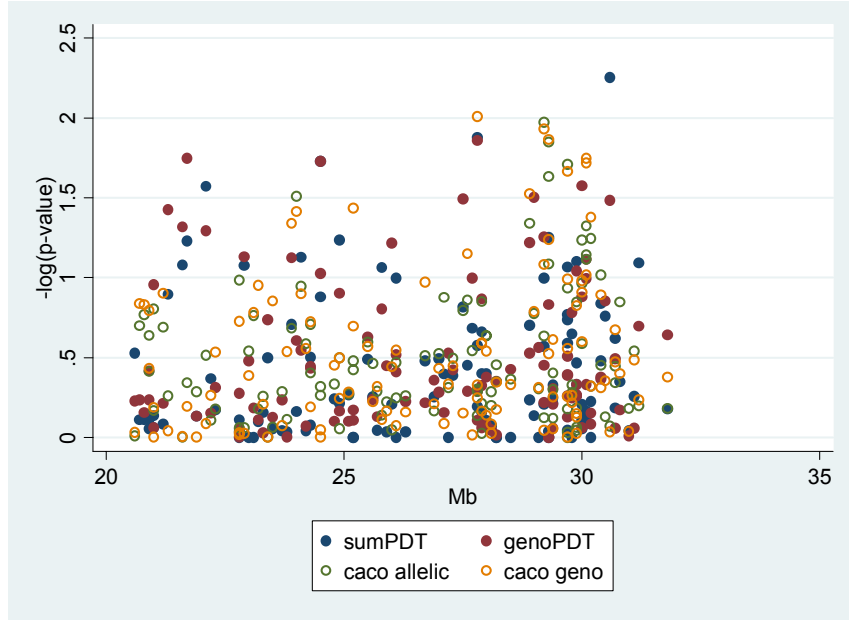


Figure 5-2. Preliminary Association Analyses on Chromosome 16. All analyses were performed in grades 345vs12.

By using a genome-wide linkage screen in the families and searching for areas of association that overlapped in both the family-based and case-control datasets, these preliminary analyses narrowed the search to a particular chromosomal segment. However, the interval of interest still spanned from 10-31 Mb and contained hundreds of genes. Our goal was to further narrow the minimum candidate region, and then to select candidate genes to test for association. To do this we increased SNP density across the region and tested for linkage and association.

Materials and Methods

Ascertainment, Genotyping, and Quality Control

Multiplex and singleton families, cases, and controls were ascertained as described previously (Chapters III and IV). Table 5-1. describes the two independent study populations.

Table 5-1. Characteristics of the Study Populations for the Screening Analyses. Mx=multiplex, Fams=families, SD=standard deviation.

| | Family Dataset | Independent Case-Control Dataset | |
|---------------|---|----------------------------------|--------------|
| | 541 phenotyped | 591 cases | 256 controls |
| | (125 Mx, 84 Singleton Fams) | (grades 3,4,5) | (grades 1,2) |
| Grade | 3: 27.9% | 3: 22.5% | 1: 74.6% |
| | 4: 15.3% | 4: 12.7% | 2: 25.4% |
| | 5: 56.8% | 5: 64.8% | |
| Mean Age, SD | 67.0, 10.3 (unaffected) 74.7, 9.1 (affected) | 76.5(7.7) | 66.4(8.2) |
| % female | 66 | 64 | 55.9 |
| % ever smoked | 57.1(affected) 49.5(unaffected) | 59.8 | 49.3 |

In the family-based dataset 330 SNPs were genotyped across the length of chromosome 16 as part of a genome-wide screen using Illumina genotyping technology. Of these 330 SNPs, 194 were within our broad region of interest from 10-31 Mb (Appendix Table A-1). In the independent case-control dataset, 149 SNPs in the region of interest were genotyped using Taqman assays from Applied Biosystems. The SNPs in the case-control dataset overlapped with those in the family-based dataset, with an average spacing of approximately 1 SNP every 66 Kb.

We verified that all SNPs were in Hardy-Weinberg equilibrium (HWE) in both the family-based and case-control datasets. HWE in the case-control dataset was examined both in the overall dataset and separately in cases and controls. We used only founders to estimate allele frequencies in the family-based dataset, except when a family did not have any founders genotyped. For those families, one individual was selected at random to contribute to the allele frequency calculation. For each SNP, 95% of the samples tested had to produce a genotype for the marker to be used in subsequent analyses.

Linkage Disequilibrium

We examined the patterns of linkage disequilibrium in the family-based and case-control datasets using Haploview. We studied the overall case-control dataset, in addition to cases and controls separately. This analysis allowed us to: 1) select a subset of markers for the multipoint linkage analysis that had pair-wise r-squared values no greater than 0.16, as markers in strong LD may bias LOD scores when parental genotypes are missing (Boyles et al. 2005) 2) study the pattern of linkage disequilibrium in the overall dataset and also separately in cases and controls, as differing patterns may indicate a region of

association and 3) determine whether those markers showing interesting results (LOD>2.0 or $p<0.01$) in subsequent analyses were independent or coming from a block of LD.

Statistical Analysis in the Family-based Dataset

Two-point dominant and recessive LOD scores were calculated with Fastlink (Cottingham, Idury, and Schaffer 1993; Schaffer et al. 1994), two-point nonparametric LOD scores were calculated with Allegro (Gudbjartsson et al. 2000; Gudbjartsson et al. 2005), and Merlin (Abecasis et al. 2002) was used for estimating multipoint nonparametric, dominant, and recessive LOD scores. We specified the following parameters for the parametric models: dominant model disease allele frequency = 0.01, $f_0=0.0000$, $f_1=0.0001$, $f_2=0.0001$, where f_i is the penetrance of an individual with i susceptibility alleles; recessive model disease allele frequency = 0.14, $f_0=0.0000$, $f_1=0.0000$, $f_2=0.0001$. We used OSA (Hauser et al. 2004) to consider linkage after accounting for CFH Y402H, LOC387715 A69S, smoking, blood pressure, IOP, and BMI separately as covariates. Families were ranked either from low to high or high to low based on their average score for the covariates (for example, the average BMI for a family), and then OSA chose the subset of families that would maximally increase the evidence for linkage in the region and calculated a new LOD score in the subset. Empirical p-values were used to assess whether the difference in the unconditional LOD score and the LOD score in a covariate-based subset was statistically significant. Single-marker association was tested in the families with APL and PDT, and two-SNP sliding window haplotypes were tested in APL. We also stratified the family dataset by

LOC387715 A69S and smoking and tested for association with APL. In these analyses, only families in which all affecteds had the same status (for example, all were smokers) were used. Individuals in a family that were missing LOC387715 A69S or smoking data were classified as having unknown AMD status. We did not stratify by CFH Y402H because there were fewer than 15 CFH Y402H non-risk families.

Statistical Analysis in the Case-control Dataset

Tests of allelic and genotypic association in the case-control dataset were performed using 2x2 and 2x3 contingency table analyses, and Haploview was used to test haplotypes in blocks defined by the Gabriel et al. method for association (Gabriel et al. 2002). Because age may be a confounder, we also tested for allelic association in an age-matched case-control dataset of 137 pairs over age 70. Stratified analyses were performed using CFH Y402H, LOC387715 A69S, a combination of CFH Y402H and LOC387715 A69S, and smoking as covariates. For the CFH Y402H/LOC387715 A69S combinations, the C allele of CFH Y402H and the T allele of LOC387715 A69S were considered risk alleles. Individuals that did not carry risk alleles at either locus were classified as low risk, individuals that carried risk alleles at one locus but not the other were classified as medium risk, and individuals that carried risk alleles at both loci were classified as high risk.

Analysis by Grade of AMD

All analyses were performed in the group of all AMD cases (grades 3, 4, and 5) compared to all AMD controls (grades 1 and 2). We also carried out a subset of these

analyses in the neovascular AMD cases only (grade 5 at least one eye or grade 5 both eyes compared to grade 1 controls), to consider the possibility that this susceptibility locus predisposes specifically to the neovascular form of AMD. These analyses included 2pt LOD score calculations and single-marker tests of association, but due to the reduced sample size, not the multipoint linkage analysis, OSA, stratified associations, or haplotype analysis.

Results

Linkage Disequilibrium

The pattern of linkage disequilibrium in the family-based dataset was similar to the pattern in the case-control dataset (Figures 5-3 and 5-4). Only a few small LD blocks were present between 10-27 Mb. Larger blocks of strong LD occurred between 27-31 Mb.

Linkage

The peak multipoint LOD score of 2.2 occurred at 22.8 Mb using a dominant model. Nonparametric and recessive LOD scores peaked at the same location (LOD=1.6 and 1.4, respectively, Figure 5-5). Several two-point LOD scores in this region exceeded 1.0, the largest being 3.1 at 24.2 Mb in the CACNG3 gene.

Using Ordered Subset Analysis (OSA) to examine the evidence for linkage after taking into account CFH Y402H, LOC387715 A69S, pack-years of cigarette smoking, blood pressure, IOP, and BMI as covariates, did not significantly change the LOD score

($p > 0.10$ for all). This implies that the evidence for linkage to this region of chromosome 16 did not increase after considering these covariates, not that there is no linkage to the region. The largest change was an increase of 1.56 LOD score units at 17.8 Mb to a max LOD of 2.56 in the “CFH Y402H low” families (Figure 5-6). Consistent with previous reports, the LOD score in the BMI high subset rose to ~ 2.3 at 22 Mb.

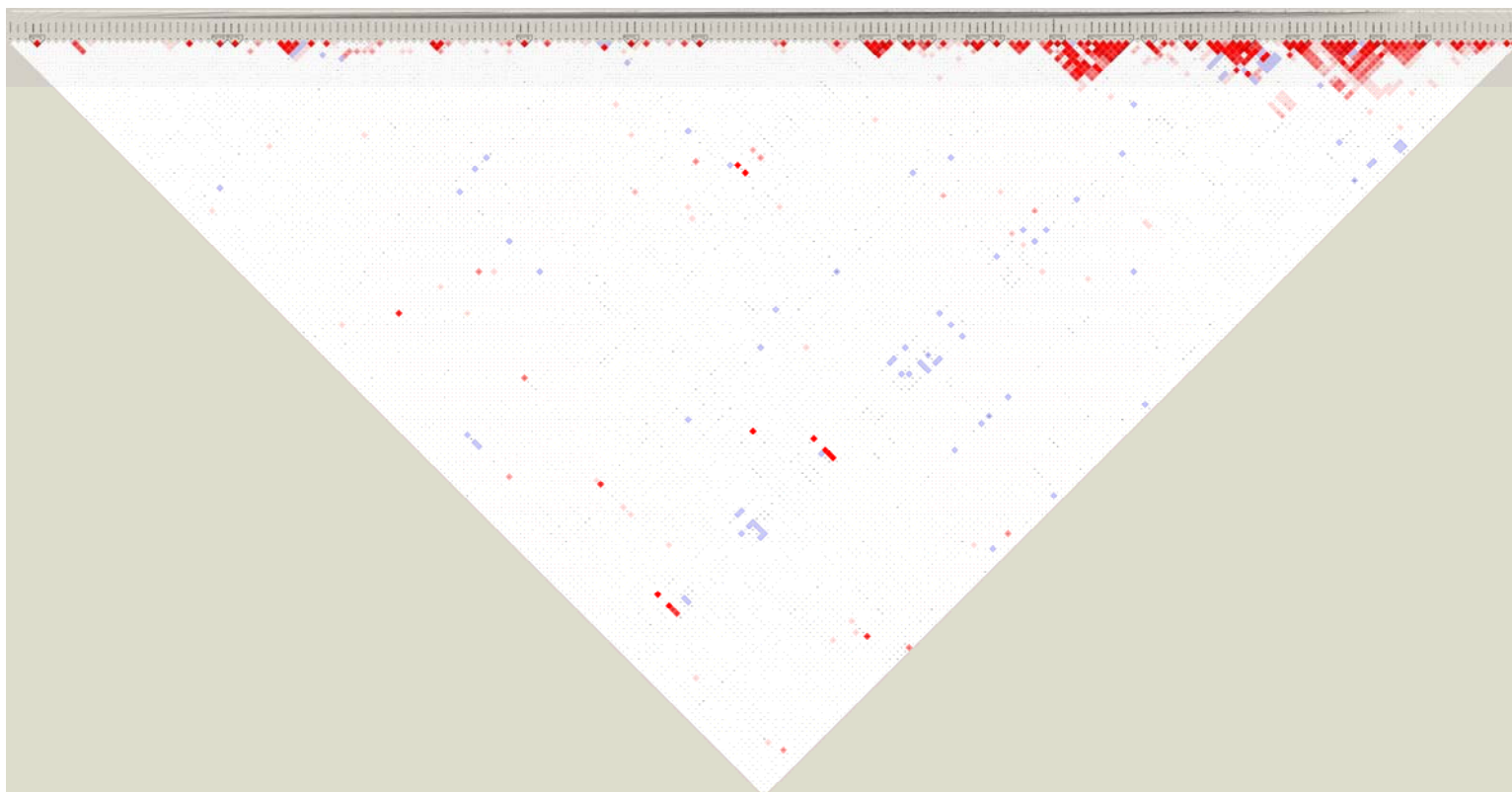


Figure 5-3. Linkage Disequilibrium in the Family-based Dataset Chr. 16p12 Grades 345vs12

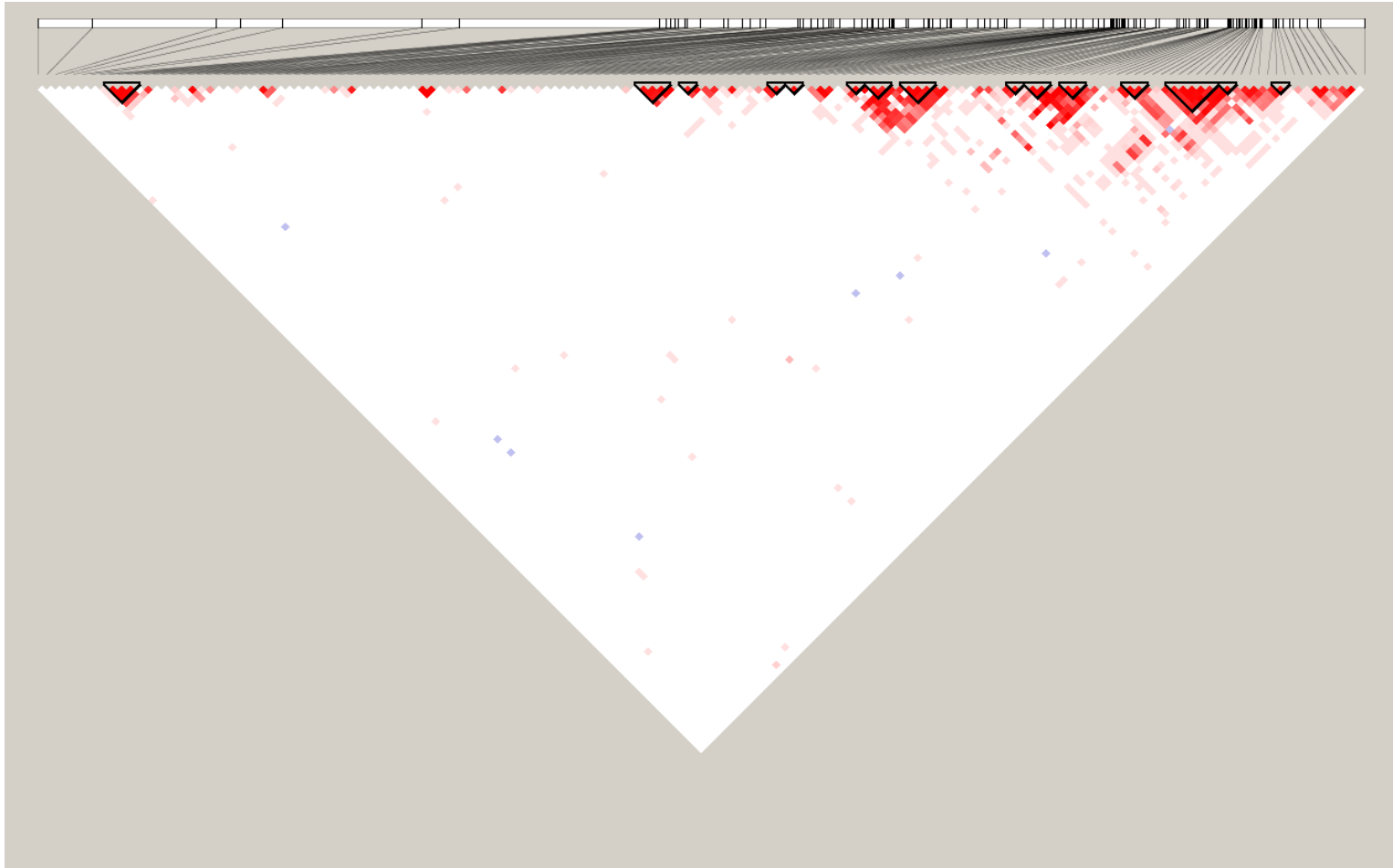


Figure 5-4. Linkage Disequilibrium in the Case-control Dataset Chr. 16p12 Grades 345vs12. Black triangles indicate haplotype blocks, as defined by the Gabriel et al. method implemented in Haploview software (Gabriel et al. 2002).

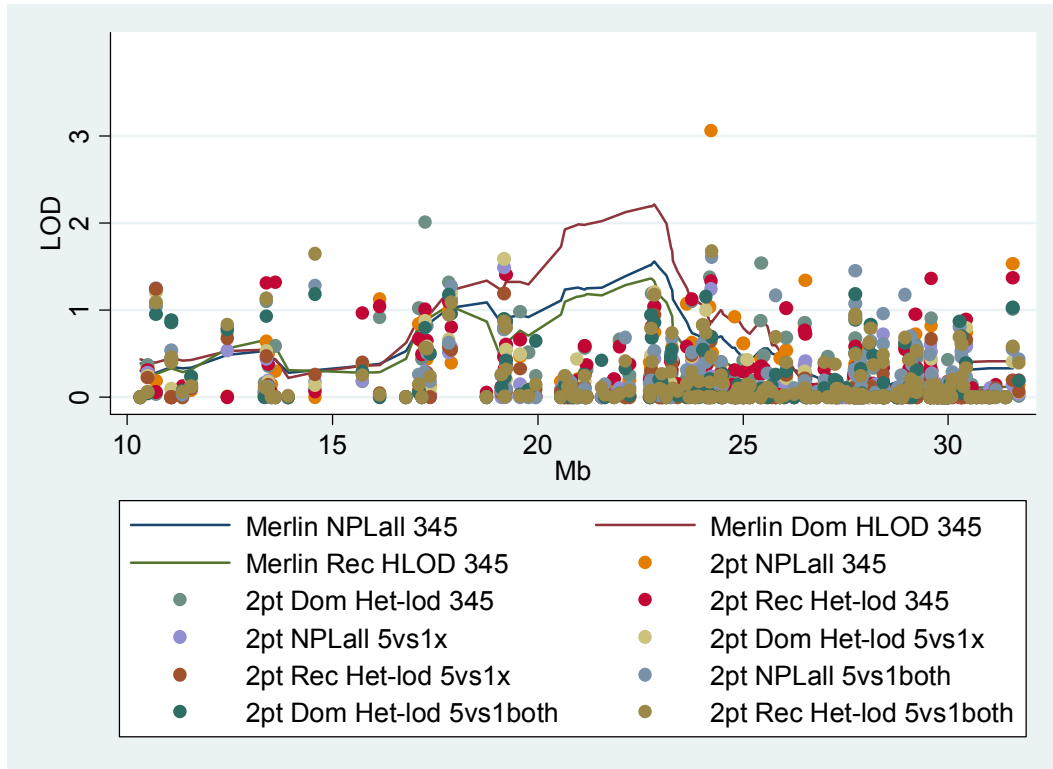
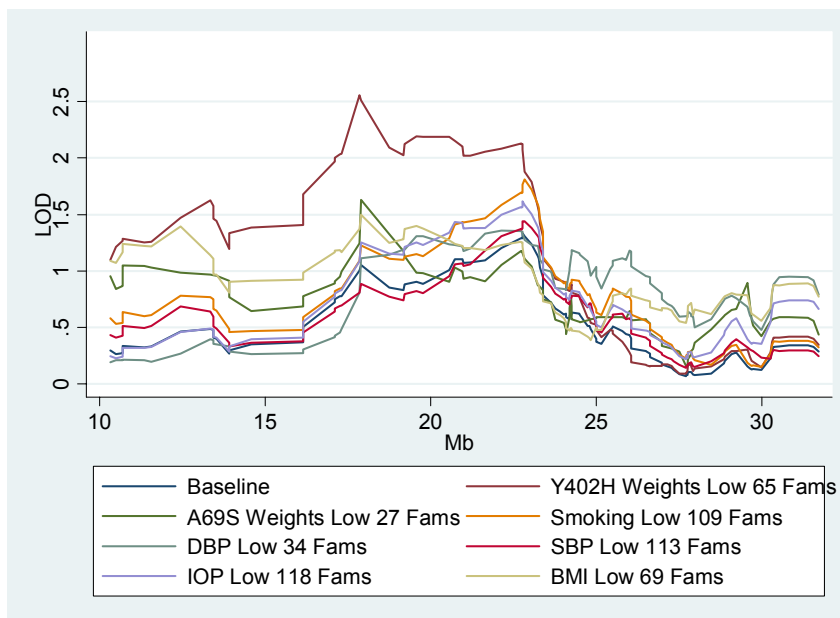


Figure 5-5. Linkage Analysis in the Region of Interest on Chr. 16p12

A. OSA on Chromosome 16p12 Low to High Covariate Rankings



B. OSA on Chromosome 16p12 High to Low Covariate Rankings

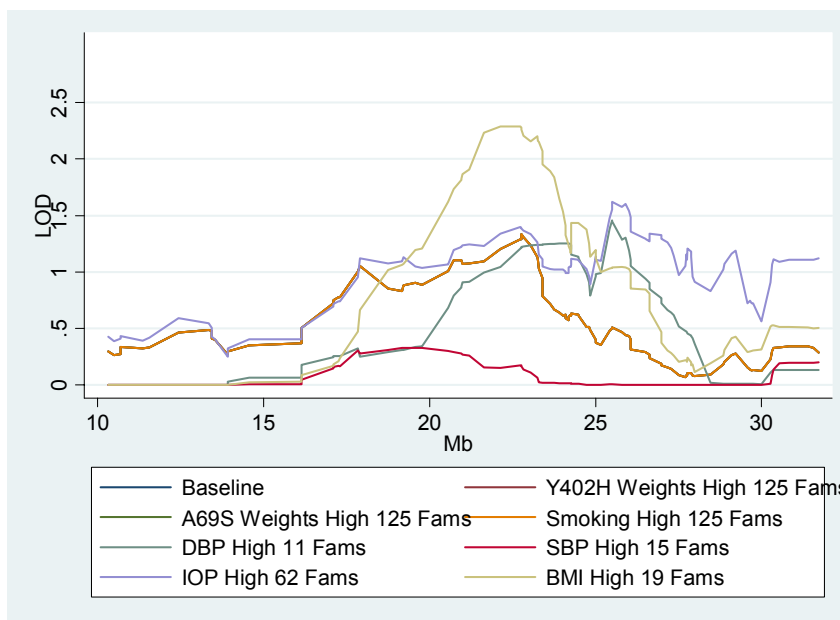


Figure 5-6. Ordered Subset Analysis on Chromosome 16p12. CFH Y402H and LOC387715 A69S covariates were weighted as the proportion of affecteds in the family carrying the risk allele. Smoking was measured in pack-years; non-smokers were coded as 0 pack-years. The total number of families used was 125. DBP=diastolic blood pressure, SBP=systolic blood pressure, IOP=intraocular pressure, BMI=body mass index. Note, none of the changes in LOD score were statistically significant ($p > 0.10$ for all).

Association

We used both the Pedigree Disequilibrium Test (PDT) and the Association in the Presence of Linkage (APL) methods to test for single-marker association in the family-based dataset. APL is more powerful than PDT in many circumstances (Martin et al. 2003), but APL is not valid in the 20 extended families that comprise nearly 10% of our dataset. Overall, PDT and APL gave similar results, with the most significant results ($p < 0.01$) clustering between 27.5 and 29 Mb (APL results, Figure 5-7, panel A). Tests of allelic and genotypic association in the case-control dataset also produced a cluster of interesting results in this area, the strongest being a SNP in IL4R in the age-matched pairs (allelic association $p < 0.001$, Figure 5-7, panel A).

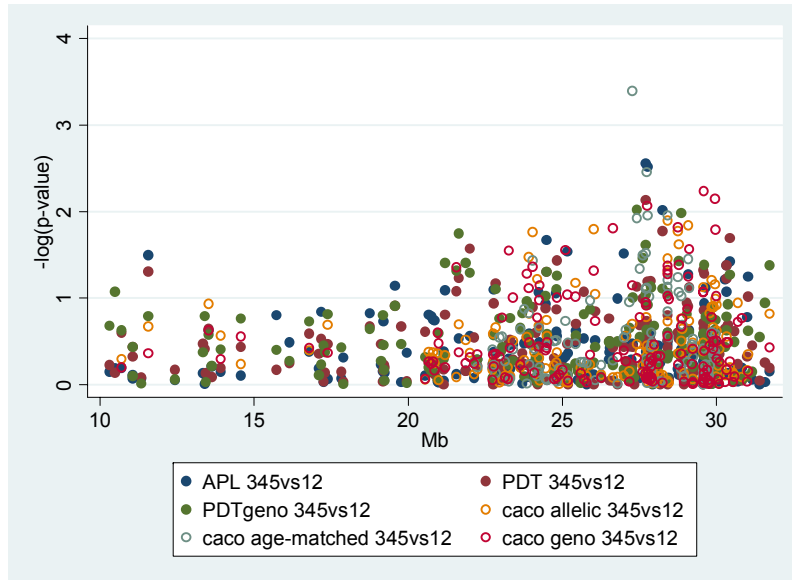
We used a two-SNP sliding window in APL to test haplotypes for association in the family-based dataset. Haplotypes in the IL4R gene at 27.2 Mb and the Q7Z6F8 gene at 27.7 Mb were strongly associated with disease risk ($p < 0.001$ and $p = 0.006$, respectively, Figure 5-7, panel B). In the case-control dataset we tested haplotype blocks defined by Haploview for association. We also “forced” Haploview to test any 2-SNP haplotype with $p < 0.01$ in the family-based dataset. A 4-SNP haplotype spanning the ATXN2L, SH2B, and CD19 genes at 28.7 Mb was associated with AMD ($p = 0.005$, Figure 5-7, panel B). Unfortunately, one of the two SNPs that produced the strongest haplotype results in the IL4R gene in the family-based dataset had not been genotyped in the case-control dataset. Using the next closest SNP available in the case-control dataset, this haplotype trended towards significance ($p = 0.06$). The haplotype in the Q7Z6F8 gene was not replicated in the case-control dataset ($p = 0.35$).

Several SNPs clustering between 23-30Mb produced interesting results in the neovascular AMD analyses in both the family-based and case-control datasets (Figure 5-7, panels C and D). These results agreed well with the analyses comparing all cases to all controls (grades 345vs12), and suggest that the chromosome 16 AMD locus contributes to overall AMD rather than a particular subtype of the disease.

Finally, we also tested for association after stratifying with regards to LOC387715 A69S and smoking in the families and on CFH Y402H, LOC387715 A69S, the combination of CFH Y402H and LOC387715 A69S risk genotypes, and smoking in the case-control dataset. Overall, the stratification did not substantially change the association results (Figure 5-8). Though there were a few strongly associated SNPs in various subsets (for example, a SNP in CACNG3 in the CFH Y402H/LOC387715 A69S combination low risk group and a SNP in CLN3 in smokers, $p < 0.001$ for both), there was no clear pattern of several nearby markers being associated in the same subset.

For completeness, Table 5-2 lists each SNP that produced a $LOD > 2.0$ or $p < 0.01$, along with its associated Build 36 Mb position, gene, and analysis method.

A. Single-marker Association Grades 345vs12



B. Haplotype Association Grades 345vs12

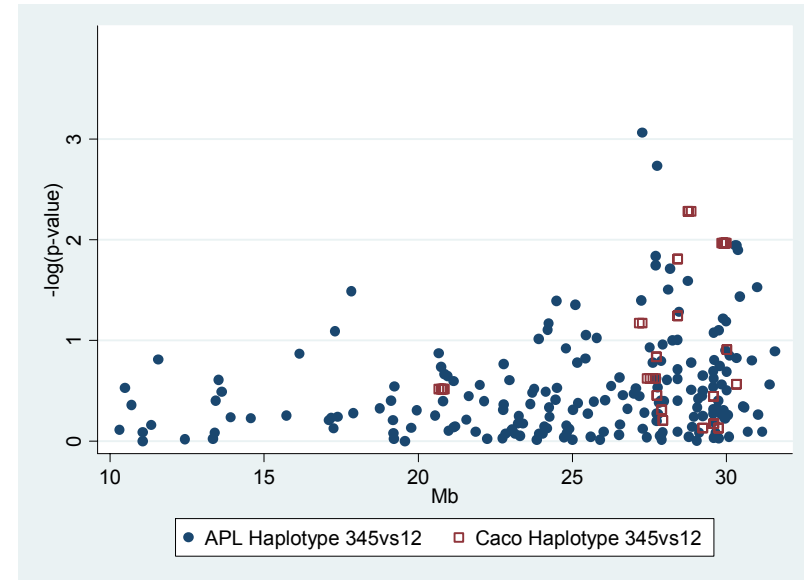
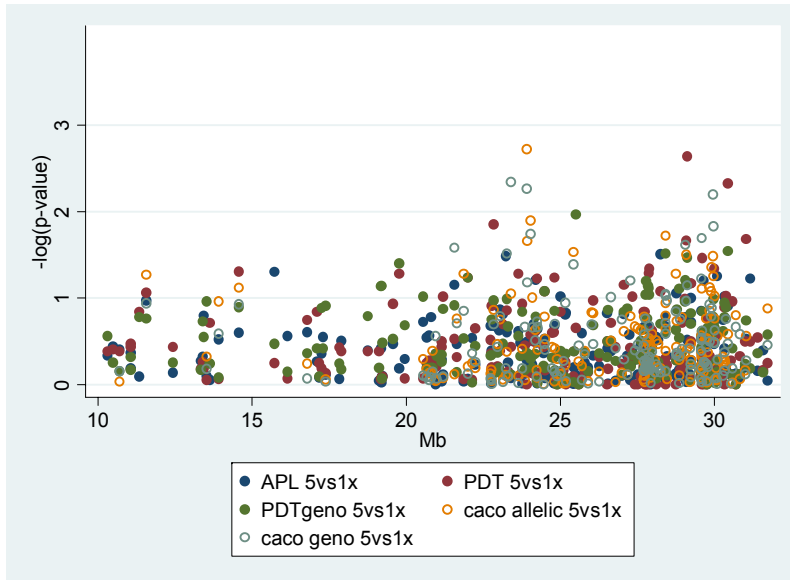


Figure 5-7. Association on Chromosome 16p12. For APL haplotypes the $-\log$ of the global p-value is plotted at the Mb location for the first SNP in the 2-SNP window. The case-control haplotypes were generated using the blocks defined by Haploview and may contain several markers. The $-\log$ of the p-value corresponding to the most strongly associated haplotype in each block is plotted.

C. Single-marker Association Grade 5vs1 at least one eye



D. Single-marker Association Grade 5vs1 both eyes

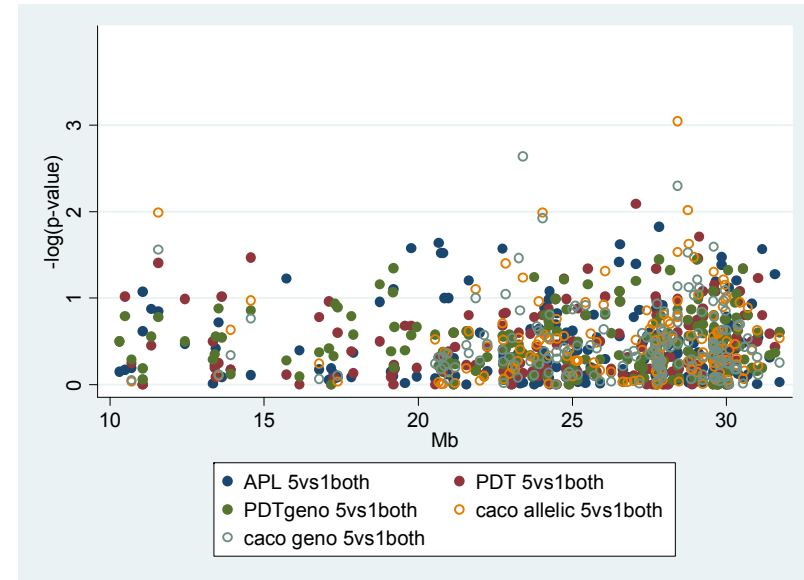


Figure 5-7 cont.

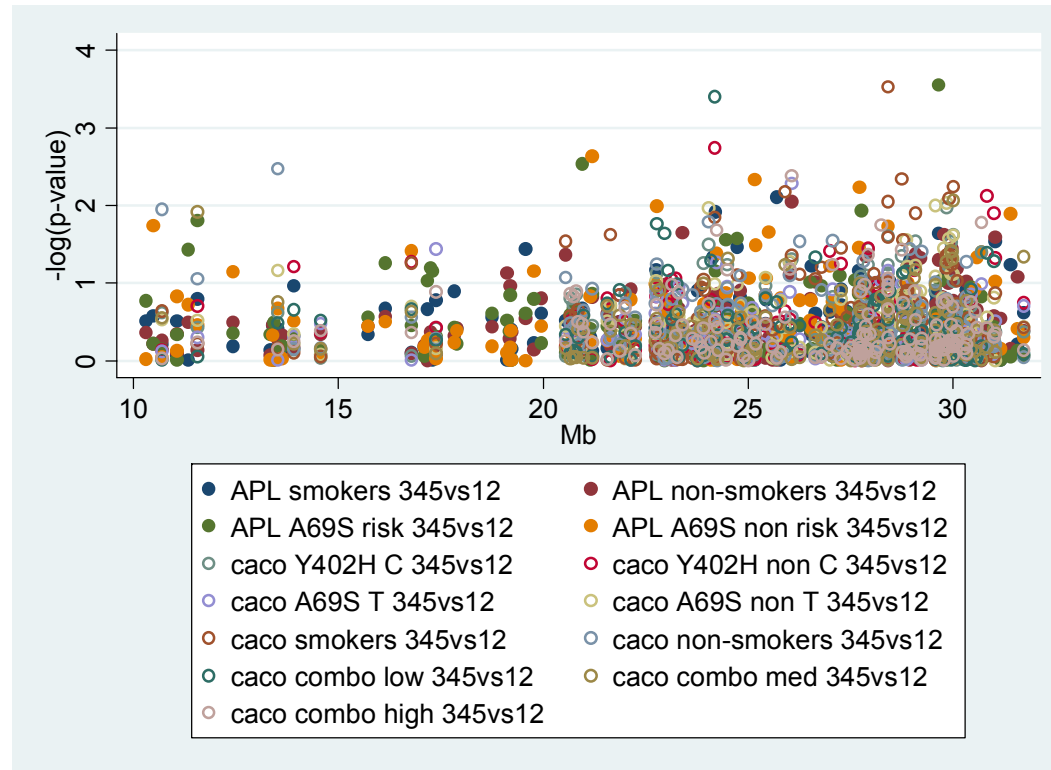


Figure 5-8. Stratified Association Analysis Grades 345vs12

Table 5-2. Interesting Results from the Screening of the Broad Region of Interest on Chr16p12. To be “interesting” a SNP had to produce a LOD>2.0 or p<0.01.

| SNP | Mb | Gene | Analysis | p-value or LOD |
|-----------|------|-------------|------------------------------------|----------------|
| RS330150 | 21.0 | DNAH3 | APL LOC387715 A69S risk 345 | 0.003 |
| RS2733910 | 21.2 | CRYM | APL LOC387715 A69S NON risk 345 | 0.002 |
| RS1055740 | 21.6 | IGSF6 | Merlin Dom HLOD 345 | 2.0 |
| RS6497580 | 22.1 | EF2K | Merlin Dom HLOD 345 | 2.1 |
| RS2239331 | 22.7 | HS3ST2 | Merlin Dom HLOD 345 | 2.2 |
| RS169660 | 22.7 | HS3ST2 | Merlin Dom HLOD 345 | 2.2 |
| RS7198577 | 22.8 | HS3ST2 | Merlin Dom HLOD 345 | 2.2 |
| RS208965 | 22.8 | HS3ST2 | Merlin Dom HLOD 345 | 2.2 |
| RS208626 | 22.8 | HS3ST2 | Merlin Dom HLOD 345 | 2.2 |
| RS1011463 | 23.0 | | Merlin Dom HLOD 345 | 2.1 |
| RS2238500 | 24.2 | CACNG3 | caco CFH Y402H non C 345 | 0.002 |
| RS2238500 | 24.2 | CACNG3 | caco strat both low 345 | 4.0E-04 |
| RS757200 | 24.2 | CACNG3 | 2pt NPLall 345 | 3.1 |
| RS2345122 | 25.2 | ZNF694 | APL LOC387715 A69S NON risk 345 | 0.005 |
| RS4520838 | 25.7 | HS3ST4 | APL smokers 345 | 0.008 |
| RS7190163 | 25.9 | HS3ST4 | caco smokers 345 | 0.007 |
| HCV504442 | 26.1 | HS3ST4 | caco LOC387715 A69S T 345 | 0.005 |
| HCV504442 | 26.1 | HS3ST4 | caco strat both high 345 | 0.004 |
| HCV504442 | 26.1 | HS3ST4 | APL NON-smokers 345 | 0.009 |
| RS3024548 | 27.3 | IL4R | APL Haplo 345 | 0.001 |
| RS3024548 | 27.3 | IL4R | caco allelic age-matched 345 | 4.0E-04 |
| RS8832 | 27.3 | IL4R | APL Haplo 345 | 0.001 |
| RS232073 | 27.4 | GTF3C1 | PDTgeno 345 | 0.010 |
| RS772859 | 27.7 | NP_056017.1 | PDTsum 345 | 0.007 |
| RS772859 | 27.7 | NP_056017.1 | APL 345 | 0.003 |
| RS755297 | 27.7 | Q7Z6F8 | APL LOC387715 A69S NON risk 345 | 0.006 |
| RS1644609 | 27.7 | Q7Z6F8 | caco allelic age-matched 345 | 0.004 |
| RS1644609 | 27.7 | Q7Z6F8 | APL Haplo 345 | 0.002 |

| SNP | Mb | Gene | Analysis | p-value or LOD |
|--------------|------|----------------|----------------------------------|----------------|
| RS713547 | 27.7 | Q7Z6F8 | APL Haplo 345 | 0.002 |
| RS713547 | 27.7 | Q7Z6F8 | caco geno 345 | 0.009 |
| RS1644618 | 27.8 | Q7Z6F8 | APL 345 | 0.003 |
| RS2726040 | 28.2 | NP_001019572.1 | APL 345 | 0.010 |
| RS151233 | 28.4 | CLN3 | caco smokers 345 | 3.0E-04 |
| APOB48R-9270 | 28.4 | NP_061160.1 | caco smokers 345 | 0.009 |
| RS12443881 | 28.7 | ATXN2L | caco smokers 345 | 0.005 |
| RS12443881 | 28.7 | ATXN2L | caco Haplo 345 | 0.005 |
| HCV105407 | 28.8 | ATXN2L | caco Haplo 345 | 0.005 |
| RS7193733 | 28.8 | SH2B | caco Haplo 345 | 0.005 |
| RS2904880 | 28.9 | CD19 | caco Haplo 345 | 0.005 |
| RS11859842 | 29.6 | | caco LOC387715 A69S non T 345 | 0.010 |
| RS1050881 | 29.6 | SPN | caco geno 345 | 0.006 |
| RS648559 | 29.6 | | APL LOC387715 A69S risk 345 | 2.8E-04 |
| RS4548895 | 29.8 | NP_859069.2 | caco LOC387715 A69S non T 345 | 0.009 |
| RS11901 | 29.9 | NP_919256.1 | caco smokers 345 | 0.008 |
| RS11150581 | 29.9 | Q96LL3 | caco geno 345 | 0.007 |
| RS8060511 | 30.0 | PPP4C | caco smokers 345 | 0.006 |
| RS8060511 | 30.0 | PPP4C | caco strat both med 345 | 0.009 |
| RS1046276 | 30.8 | CTF1 | caco CFH Y402H non C 345 | 0.008 |
| RS4968008 | 23.4 | GGA2 | caco geno 5vs1x | 0.005 |
| RS916677 | 23.9 | PRKCB1 | caco allelic all 5vs1x | 0.002 |
| RS916677 | 23.9 | PRKCB1 | caco geno 5vs1x | 0.005 |
| RS252313 | 29.1 | | PDTsum 5vs1x | 0.002 |
| RS11150581 | 29.9 | Q96LL3 | caco geno 5vs1x | 0.006 |
| RS2230433 | 30.4 | ITGAL | PDTsum 5vs1x | 0.005 |
| RS4968008 | 23.4 | GGA2 | caco geno 5vs1both | 0.002 |
| RS4787929 | 27.1 | | PDTsum 5vs1both | 0.008 |
| RS151233 | 28.4 | CLN3 | caco allelic all 5vs1both | 0.001 |
| RS151233 | 28.4 | CLN3 | caco geno 5vs1both | 0.005 |
| RS12443881 | 28.7 | Q8WWM7-7 | caco allelic all 5vs1both | 0.010 |

Discussion

After increasing the SNP coverage on chromosome 16, the evidence for linkage increased with peak LOD scores exceeding 2.0 at ~22 Mb. Though the OSA analysis did not show any significant difference in the LOD scores, families with a lower frequency of the CFH Y402H risk allele and higher BMI tended to be more strongly linked to chromosome 16, in keeping with the results from our initial analyses (CFH Y402H max LOD 2.56 at ~18 Mb, $p=0.11$, BMI max LOD 2.3 at 22 Mb, $p=0.21$). It is important to remember that OSA p-values are based on testing for a significant difference between the unconditional LOD score and the LOD score in a subset of the families determined by ranking on a covariate score and not on the value of the conditional LOD score vs. absence of linkage. Therefore, even though there was no “significant difference” in the LOD scores, the fact that the scores were greater than 2.0 overall and in subsets of the families is suggestive of linkage to this region.

Interestingly, many of the most significant associations in both datasets were clustered at ~24-30 Mb, farther downstream of the multipoint linkage peak but still within a region with many 2pt LOD scores exceeding 1.0. In agreement with the OSA results in the CFH Y402H low subset, CACNG3, the gene with the highest 2pt LOD score, also contained a SNP that was associated in the CFH Y402H non-risk allele carriers and the CFH Y402H/LOC387715 A69S low risk combination ($p<0.002$ for both groups). SNPs in both IL4R and Q7Z6F8 were also associated with AMD in both datasets (IL4R rs3024548-rs8832 APL haplotype $p=0.001$, rs3024548 age-matched pairs allelic association $p<0.0001$; Q7Z6F8 rs1644609 APL $p=0.003$, age-matched pairs allelic association $p=0.004$). In agreement with the overall analyses, the most significant associations in the

neovascular AMD analyses also clustered between 27-30 Mb, indicating that the chromosome 16p12 locus may predispose to the broad AMD phenotype.

A major strength of our study, compared to others focused on the genetics of AMD, is the wealth of clinical and lifestyle data we have ascertained for each of our participants. Since AMD exhibits both phenotypic variability and locus heterogeneity, we can use information on smoking history, BMI, IOP, and genotypes at CFH Y402H, LOC387715 A69S, and R32Q to create more homogeneous subsets, as was done in the OSA and stratified association analyses. Unfortunately, testing for linkage and association in many subgroups also increases the multiple testing problem. To balance this tradeoff, when selecting candidate genes for follow-up analyses, we will give special weight to genes with positive results in both the family-based and case-control datasets.

In summary, based on both the linkage and association analyses, we more than halved the size of the minimum candidate region from 21 million basepairs down to approximately 9 Mb at chr16:21-30 Mb. Because the marker density is higher from 21-30 Mb compared to 10-20 Mb, it is possible that higher number of interesting findings in this region includes spurious results caused by multiple testing. Therefore, the convergence of analytical methods, gene expression data, and known biology of the genes within this interval will be necessary for selecting candidate genes for further study.

CHAPTER VI

SELECTION AND TESTING CANDIDATE GENES ON CHROMOSOME 16P12 FOR LINKAGE AND ASSOCIATION WITH AMD

Introduction

AMD is the leading cause of vision loss in the elderly. CFH Y402H, LOC387715 A69S, and CFB R32Q are confirmed AMD susceptibility loci, and environmental factors like cigarette smoking and increased body mass index also increase risk for AMD (Chapter I).

Chromosome 16p12 has been consistently linked with an AMD susceptibility locus (Chapter V). We fine-mapped this region by increasing SNP density and testing for linkage and association. To further localize this AMD susceptibility gene, we used the process of genomic convergence (Hauser et al. 2003), whereby gene expression data and screening analysis results are collated and genes with interesting results from both methods are selected as candidates.

Materials and Methods

Selection of Candidate Genes

To aid in prioritization of candidate genes, we combined the available information on each gene from multiple databases with results from our own analysis (Chapter V). For each of the genes that reside between 21-30 Mb we noted the location of the gene,

potential function, number and type of SNPs, serial analysis of gene expression (SAGE) data, and analysis results.

Of the 29 genes with at least one interesting genetic result (according to the arbitrarily chosen cut-offs of $\text{LOD} > 2.0$ or $p < 0.01$), 12 were classified as interesting using multiple (2 or more) methods (Figure 6-1). All of these genes were expressed in the eye (based on a SAGE tag count > 5) or lacked SAGE data, so expression information was not useful for narrowing this list. Of these 12 genes, only 4 had interesting results in both the family-based and case-control datasets. These were CACNG3 (24.1 Mb), HS3ST4 (25.6-26.0 Mb), IL4R (27.2 Mb) and Q7Z6F8 (27.7-28.0 Mb).

CACNG3 codes for the gamma subunit of L-type voltage-dependent calcium channels. This gene is highly expressed in the retina and neural tissues, compared to the rest of the body (National Eye Institute Serial Analysis of Gene Expression data). It was recently shown that retinal pigment epithelial cells require L-type Ca^{2+} channels to properly generate a light peak and that mutations in the gene responsible for Best's disease, another macular degenerative disorder, affect L-type channel activation kinetics and voltage dependence (Rosenthal et al. 2006). Furthermore, use of calcium channel blockers was weakly associated with AMD in the Beaver Dam Eye Study and the Women's Health Initiative Sight Exam Ancillary Study (Klein et al. 2001b; Klein et al. 2007). In the screening analyses for chromosome 16 (Chapter V, Table 5-2), CACNG3 was both linked to disease in the family-based dataset and associated in the case-control dataset (2pt LOD 3.1, allelic association p-value 0.0004 in individuals who do not carry the risk alleles at either CFH Y402H or LOC387715).

All Interesting Results

ALDOA
CACNG3
CD19
CLN3
CRYM
CTF1
DNAH3
EF2K
GGA2
GTF3C1
HS3ST2
HS3ST4
IGSF6
IL4R
ITGAL
KCTD14
NP_001019572.1
NP_056017.1
NP_061160.1 (APOB48R)
NP_859069.2
NP_919256.1
PPPC4
PRKCB1
Q7Z6F8
Q8WWM7-7 (ATXN2-like)
Q96LL3
Q9NRF3
SPN
ZNF694

Multiple Methods

CACNG3
CLN3
HS3ST2
HS3ST4
IL4R
NP_056017.1
NP_061160.1 (APOB48R)
NP_919256.1
PPP4C
PRKCB1
Q7Z6F8
Q8WWM7-7 (ATXN2-like)
Q96LL3

Expressed in Eye

CACNG3
HS3ST2
HS3ST4
IL4R
PPP4C
PRKCB1

No SAGE Data
CLN3
NP_056017.1
NP_061160.1 (APOB48R)
NP_919256.1
Q7Z6F8
Q8WWM7-7 (ATXN2-like)
Q96LL3

Both Datasets

IL4R
CACNG3
HS3ST4
Q7Z6F8

Figure 6-1. Selection of Candidate Genes. Interesting results were defined as a $\text{LOD} > 2.0$ or $p < 0.01$. “Both Datasets” refers to the family-based and case-control datasets.

HS3ST4 encodes the enzyme that generates 3-O-sulfated glucosaminyl residues on heparan sulfate. CFH binds both heparin and heparan sulfate on endothelial cell surfaces, thereby preventing attack of self-tissues by the immune system (Pangburn, Atkinson, and Meri 1991; Jokiranta et al. 2005). Since polymorphisms in CFH are associated with AMD, it is possible that variants in HS3ST4 may interfere with proper production of heparan sulfate, leading to absence of binding by CFH and increased risk of AMD. SNPs

in this gene were associated with AMD in both the family-based and case-control datasets ($p < 0.01$), but only when stratifying by smoking, LOC387715 A69S, and the CFH Y402H/LOC387715 A69S high risk combination.

IL4R is expressed in the RPE and regulates IgE production. Large amounts of IgE were found in the connective stroma of the subretinal membrane and in the new blood vessel walls of AMD patients (Baudouin et al. 1992). A 2-SNP haplotype in IL4R was strongly associated with AMD in the family-based dataset ($p < 0.001$), and one of the SNPs from this haplotype was associated in the age-matched case-control dataset of 137 pairs ($p < 0.001$).

Very little is known about the function of Q7Z6F8, also known as GSG1L for “germ cell associated 1-like”. However, due to the screening nature of our approach, we were able to identify novel candidates without relying on known biology. Both single SNPs and 2-SNP haplotypes in this gene were associated with AMD in the families, and a genotype at one of these SNPs was related to disease status in the case-control dataset.

Therefore, after compiling the screening results of this region, gene expression information, and known biology of the genes, we selected 4 candidates for AMD: CACNG3, HS3ST4, IL4R, and Q7Z6F8 (Table 6-1).

Table 6-1. Characteristics of the 4 Candidate Genes. All genes, except Q7Z6F8, are known to be expressed in the eye (NEI SAGE Data). None of the genes are known to be specifically rod-, cone-, or RPE-associated. CACNG3 and HS3ST4 have one known transcript, and IL4R and Q7Z6F8 each have 3 transcripts. Biotype, Status, and SNP counts from Ensembl. Start position and size from NCBI Build 36.

| Gene | Description | Start Position (bp) | Size (bp) | Biotype | Status | Total # SNPs | Non-synonymous | Synonymous | # Genotyped in Screen |
|--------|--|---------------------|-----------|----------------|--------|--------------|----------------|------------|-----------------------|
| CACNG3 | Voltage-dependent calcium channel gamma-3 subunit | 24174382 | 106856 | protein coding | Known | 344 | 0 | 1 | 5 |
| HS3ST4 | Heparan sulfate glucosamine 3-O-sulfotransferase 4 | 25611240 | 443830 | protein coding | Known | 1523 | 2 | 0 | 3 |
| IL4R | Interleukin-4 receptor alpha chain precursor | 27259005 | 24594 | protein coding | Known | 258 | 17 | 13 | 2 |
| Q7Z6F8 | | 27706357 | 275966 | protein coding | Known | 1045 | 0 | 0 | 14 |

Study Populations

The family-based and case-control samples for these analyses overlap with the samples used previously (Chapters V). Approximately 19 multiplex families, 110 cases, and 30 controls were added to the datasets (Table 6-2). Ascertainment, grading, DNA extraction, and quality control measures for the new samples were performed according to the same protocols used previously. Minor allele frequency, HWE p-value, and genotyping efficiency for each SNP are reported in Table 6-3.

Table 6-2. Characteristics of the Study Populations. Mx=multiplex

| | Family Dataset | Independent Case-Control Dataset | |
|-------------------|------------------------------------|---|--------------|
| Total Individuals | 559 phenotyped | 701 cases | 286 controls |
| | (144 Mx, 79 Singleton families) | (grades 3,4,5) | (grades 1,2) |
| Grade | 3: 29.3% | 3: 26.7% | 1: 72.0% |
| | 4: 14.4% | 4: 12.6% | 2: 28.0% |
| | 5: 56.3% | 5: 60.7% | |
| Mean Age (sd) | 72.8(9.9) | 76.5(7.7) | 66.9(8.4) |
| % female | 66.5 | 63.6 | 55.6 |
| % ever smoked | 56.3(affected) 46.9(unaffected) | 60.9 | 48.7 |

Table 6-3. Description of the SNPs Analyzed. Chr=chromosome, Geno Eff= genotyping efficiency, HWE= Hardy-Weinberg Equilibrium, MAF=minor allele frequency

| SNP | Gene | Chr | Position | Minor Allele | Geno Eff (%) | Family 345vs12 | | Case-Control 345vs12 | | |
|------------|-----------|-----|-----------|--------------|--------------|----------------|------|----------------------|-----------|--------------|
| | | | | | | HWE All | MAF | HWE All | HWE Cases | HWE Controls |
| RS1061170 | CFH | 1 | 194925860 | T | 99 | 0.68 | 0.41 | 0.98 | 0.90 | 0.12 |
| RS9332739 | CC2 | 6 | 32011783 | C | 97 | 0.08 | 0.03 | 1.00 | 0.76 | 1.00 |
| RS641153 | CFB | 6 | 32022159 | A | 99 | 1.00 | 0.06 | 0.52 | 1.00 | 0.48 |
| RS10490924 | LOC387715 | 10 | 124204438 | T | 99 | 1.00 | 0.43 | 5E-04 | 2.5E-03 | 0.80 |
| RS2238498 | CACNG3 | 16 | 24178560 | A | 98 | 0.75 | 0.22 | 0.59 | 0.19 | 0.33 |
| RS2238500 | CACNG3 | 16 | 24192709 | G | 100 | 0.24 | 0.44 | 0.73 | 0.59 | 0.88 |
| RS991911 | CACNG3 | 16 | 24198554 | A | 99 | 0.20 | 0.38 | 0.77 | 0.65 | 0.96 |
| RS2283551 | CACNG3 | 16 | 24200786 | C | 98 | 0.17 | 0.21 | 0.43 | 0.46 | 0.95 |
| RS9921732 | CACNG3 | 16 | 24204971 | G | 98 | 1.00 | 0.02 | 0.80 | 0.53 | 1.00 |
| RS739747 | CACNG3 | 16 | 24207067 | T | 98 | 0.71 | 0.33 | 0.97 | 1.00 | 0.82 |
| RS11640935 | CACNG3 | 16 | 24211075 | T | 98 | 0.97 | 0.27 | 0.99 | 0.50 | 0.34 |
| RS9926669 | CACNG3 | 16 | 24225803 | A | 96 | 0.72 | 0.20 | 0.06 | 0.17 | 0.26 |
| RS757200 | CACNG3 | 16 | 24227195 | T | 98 | 0.43 | 0.30 | 0.06 | 0.06 | 0.68 |
| RS1859200 | CACNG3 | 16 | 24228292 | G | 98 | 0.40 | 0.45 | 1.00 | 0.56 | 0.38 |
| RS7187560 | CACNG3 | 16 | 24229868 | T | 95 | 0.77 | 0.46 | 0.74 | 0.53 | 0.07 |
| RS4787433 | CACNG3 | 16 | 24230933 | C | 98 | 0.77 | 0.38 | 0.60 | 0.95 | 0.20 |
| RS8051597 | CACNG3 | 16 | 24241029 | C | 99 | 0.60 | 0.24 | 0.72 | 0.74 | 1.00 |
| RS11640437 | CACNG3 | 16 | 24245904 | G | 99 | 0.21 | 0.26 | 0.22 | 0.32 | 0.54 |
| RS2238518 | CACNG3 | 16 | 24247639 | C | 97 | 0.91 | 0.43 | 0.51 | 0.19 | 0.48 |
| RS2238521 | CACNG3 | 16 | 24256092 | A | 98 | 0.85 | 0.25 | 0.51 | 0.92 | 0.35 |
| RS2189290 | CACNG3 | 16 | 24264034 | C | 100 | 0.13 | 0.36 | 0.24 | 1.00 | 0.02 |
| RS8048828 | CACNG3 | 16 | 24266934 | G | 98 | 0.05 | 0.45 | 0.39 | 0.59 | 0.01 |
| RS9925471 | CACNG3 | 16 | 24267711 | G | 99 | 0.98 | 0.04 | 1.00 | 1.00 | 1.00 |
| RS12928078 | CACNG3 | 16 | 24280658 | A | 98 | 0.80 | 0.05 | 0.04 | 0.03 | 0.90 |
| RS8049719 | HS3ST4 | 16 | 25680705 | A | 98 | 1.00 | 0.10 | 0.91 | 1.00 | 0.96 |
| RS7197707 | HS3ST4 | 16 | 25682282 | A | 98 | 0.04 | 0.45 | 0.80 | 0.56 | 0.12 |
| RS11074721 | HS3ST4 | 16 | 25692399 | T | 98 | 0.09 | 0.29 | 1.00 | 0.72 | 0.61 |

Table 6-3. cont.

| SNP | Gene | Chr | Position | Minor Allele | Geno Eff (%) | Family 345vs12 | | Case-Control 345vs12 | | |
|------------|--------|-----|----------|--------------|--------------|----------------|------|----------------------|-----------|--------------|
| | | | | | | HWE All | MAF | HWE All | HWE Cases | HWE Controls |
| RS8047773 | HS3ST4 | 16 | 25735159 | G | 99 | 0.31 | 0.24 | 1.00 | 1.00 | 1.00 |
| RS4523929 | HS3ST4 | 16 | 25772018 | T | 98 | 1.00 | 0.34 | 0.97 | 1.00 | 1.00 |
| RS6497896 | HS3ST4 | 16 | 25781487 | G | 98 | 0.51 | 0.12 | 0.02 | 0.03 | 0.62 |
| RS34215460 | HS3ST4 | 16 | 25795966 | C | 100 | 0.87 | 0.33 | 1.00 | 0.66 | 0.42 |
| RS11862736 | HS3ST4 | 16 | 25811211 | C | 99 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| RS8063351 | HS3ST4 | 16 | 25897491 | T | 99 | 0.60 | 0.05 | 0.84 | 0.96 | 0.34 |
| RS7190163 | HS3ST4 | 16 | 25907847 | C | 100 | 0.32 | 0.30 | 0.84 | 0.93 | 0.51 |
| RS7190703 | HS3ST4 | 16 | 25935944 | C | 98 | 0.75 | 0.42 | 0.24 | 0.78 | 0.09 |
| RS7188016 | HS3ST4 | 16 | 26007978 | T | 98 | 1.8E-05 | 0.50 | 0.02 | 0.04 | 0.37 |
| RS4390598 | HS3ST4 | 16 | 26010569 | C | 98 | 0.50 | 0.46 | 0.03 | 0.08 | 0.24 |
| RS6498012 | IL4R | 16 | 27239475 | C | 97 | 0.41 | 0.40 | 0.07 | 0.01 | 0.74 |
| RS3024547 | IL4R | 16 | 27261862 | T | 98 | 0.40 | 0.15 | 0.33 | 0.49 | 0.59 |
| RS3024548 | IL4R | 16 | 27262032 | G | 96 | 0.82 | 0.44 | 0.94 | 0.45 | 0.13 |
| RS2239349 | IL4R | 16 | 27266389 | A | 97 | 1.00 | 0.11 | 0.79 | 0.25 | 0.28 |
| RS2239347 | IL4R | 16 | 27266522 | C | 97 | 0.84 | 0.45 | 0.49 | 0.06 | 0.10 |
| RS3024585 | IL4R | 16 | 27267345 | A | 98 | 0.65 | 0.47 | 0.82 | 0.55 | 0.72 |
| RS3024623 | IL4R | 16 | 27272969 | T | 99 | 0.02 | 0.09 | 0.86 | 1.00 | 0.98 |
| RS2234897 | IL4R | 16 | 27281113 | C | 95 | 0.33 | 0.02 | 0.41 | 1.00 | 0.13 |
| RS1805011 | IL4R | 16 | 27281373 | C | 98 | 0.83 | 0.10 | 1.00 | 0.88 | 0.91 |
| RS2234898 | IL4R | 16 | 27281416 | T | 95 | 0.08 | 0.11 | 0.90 | 1.00 | 0.58 |
| RS1805013 | IL4R | 16 | 27281481 | T | 99 | 0.69 | 0.05 | 0.21 | 0.11 | 1.00 |
| RS1805015 | IL4R | 16 | 27281681 | C | 95 | 0.42 | 0.16 | 0.23 | 0.71 | 0.14 |
| RS8832 | IL4R | 16 | 27283288 | A | 98 | 1.00 | 0.46 | 0.02 | 0.25 | 0.01 |
| RS772859 | Q7Z6F8 | 16 | 27701027 | G | 99 | 0.26 | 0.38 | 0.81 | 0.98 | 0.54 |
| RS755297 | Q7Z6F8 | 16 | 27713895 | C | 97 | 0.82 | 0.18 | 0.71 | 0.75 | 1.00 |
| RS1559167 | Q7Z6F8 | 16 | 27726752 | T | 97 | 0.08 | 0.43 | 0.35 | 0.35 | 0.85 |
| RS1644609 | Q7Z6F8 | 16 | 27734413 | A | 100 | 0.37 | 0.40 | 0.51 | 0.69 | 0.64 |
| RS713547 | Q7Z6F8 | 16 | 27739590 | C | 99 | 0.45 | 0.15 | 0.78 | 0.66 | 0.17 |

Table 6-3. cont.

| SNP | Gene | Chr | Position | Minor Allele | Geno Eff (%) | Family 345vs12 | | Case-Control 345vs12 | | |
|------------|--------|-----|----------|--------------|--------------|----------------|------|----------------------|-----------|--------------|
| | | | | | | HWE All | MAF | HWE All | HWE Cases | HWE Controls |
| RS772858 | Q7Z6F8 | 16 | 27741798 | G | 95 | 1.00 | 0.46 | 0.41 | 0.76 | 0.37 |
| RS11645119 | Q7Z6F8 | 16 | 27742183 | T | 98 | 0.05 | 0.08 | 0.43 | 0.49 | 0.89 |
| RS705928 | Q7Z6F8 | 16 | 27749116 | G | 98 | 0.58 | 0.36 | 0.81 | 1.00 | 0.76 |
| RS1644618 | Q7Z6F8 | 16 | 27757739 | T | 97 | 1.00 | 0.28 | 0.29 | 0.28 | 0.94 |
| RS9933050 | Q7Z6F8 | 16 | 27759775 | A | 99 | 0.40 | 0.34 | 1.00 | 1.00 | 0.91 |
| RS8046590 | Q7Z6F8 | 16 | 27762807 | A | 98 | 0.77 | 0.31 | 0.63 | 0.42 | 0.82 |
| RS10438534 | Q7Z6F8 | 16 | 27772260 | A | 99 | 0.35 | 0.09 | 0.81 | 0.68 | 1.00 |
| RS6498040 | Q7Z6F8 | 16 | 27776664 | T | 98 | 0.89 | 0.40 | 0.24 | 0.25 | 0.81 |
| RS734432 | Q7Z6F8 | 16 | 27794674 | A | 98 | 0.40 | 0.35 | 0.37 | 0.56 | 0.54 |
| RS1644582 | Q7Z6F8 | 16 | 27800656 | A | 98 | 0.24 | 0.42 | 0.37 | 0.74 | 0.29 |
| RS12926773 | Q7Z6F8 | 16 | 27804621 | T | 99 | 0.16 | 0.09 | 0.07 | 0.04 | 1.00 |
| RS4788003 | Q7Z6F8 | 16 | 27806901 | T | 98 | 0.83 | 0.35 | 0.45 | 0.50 | 0.85 |
| RS2385008 | Q7Z6F8 | 16 | 27841534 | A | 100 | 0.74 | 0.34 | 0.34 | 0.21 | 0.94 |
| RS7194904 | Q7Z6F8 | 16 | 27874886 | A | 98 | 0.51 | 0.47 | 0.55 | 0.37 | 0.87 |
| RS4788017 | Q7Z6F8 | 16 | 27884330 | G | 100 | 0.02 | 0.24 | 0.78 | 0.91 | 0.38 |
| RS9941112 | Q7Z6F8 | 16 | 27885980 | A | 100 | 0.65 | 0.37 | 0.25 | 0.38 | 0.51 |
| RS1008409 | Q7Z6F8 | 16 | 27901325 | C | 98 | 0.90 | 0.33 | 0.82 | 0.99 | 0.57 |
| RS11074888 | Q7Z6F8 | 16 | 27930557 | C | 98 | 0.42 | 0.40 | 0.49 | 0.31 | 0.86 |
| RS1476507 | Q7Z6F8 | 16 | 27932571 | C | 100 | 0.24 | 0.39 | 0.37 | 0.30 | 1.00 |
| RS205418 | Q7Z6F8 | 16 | 27980171 | G | 100 | 0.59 | 0.24 | 0.01 | 0.03 | 0.32 |

SNP Selection and Genotyping

Using the Tagger algorithm in Haploview software and genotypes for SNPs in the HapMap Project, we selected ~10 additional SNPs per gene designed to capture the largest percentage of common variation in each gene. This gave us excellent coverage of IL4R, moderate coverage of CACNG3 and Q7Z6F8, and some coverage of HS3ST4. (Table 6-4). Because the SNPs were chosen based on LD patterns to be the most informative (implying a high minor allele frequency), many of the SNPs were intronic. Of the 73 total SNPs genotyped in the candidate genes, 41 were intronic, 3 were non-synonymous coding changes, 3 were synonymous coding changes, and 3 were changes in either the 5' or 3' UTR. The remaining polymorphisms were not classified. All SNPs were genotyped using Taqman Assays on Demand or Assays by Design from Applied Biosystems.

Statistical Analyses in the Family-based Dataset

We calculated multipoint nonparametric, dominant, and recessive LOD scores in the multiplex families using Merlin software. To avoid false inflation of the LOD score, we used only SNPs that were not in strong LD in this analysis ($r^2 < 0.16$ between all SNPs). Allegro was used to calculate two-point nonparametric LOD scores and Fastlink was used to calculate two-point dominant and recessive LOD scores. As before, we used OSA to examine the evidence for linkage to chromosome 16 after considering CFH Y402H, LOC387715 A69S, smoking, blood pressure, IOP, and BMI as covariates. Additionally, we included CFB R32Q.

Table 6-4. Estimated Percentage of Alleles Captured Using SNPs Previously Genotyped Compared to Percentage of Alleles Captured by Adding 10 SNPs.

| Gene | # Previously Genotyped | 0 Additional SNPs % captured | mean r^2 of captured | 10 Additional SNPs % captured | mean r^2 of captured |
|-------------|-------------------------------|---|--|--|--|
| CACNG3 | 5 | 0.11 | 0.871 | 0.60 | 0.883 |
| IL4R | 2 | 0.17 | 0.882 | 0.95 | 0.877 |
| HS3ST4 | 3 (only 2 in Hapmap) | 0.01 | 0.928 | 0.24 | 0.861 |
| Q7Z6F8 | 14 (only 4 in Hapmap) | 0.13 | 0.854 | 0.76 | 0.878 |

We tested for single-SNP association using both APL and PDT. Rather than stratify the family dataset by LOC387715 A69S and smoking, as was done previously, we instead selected the families included in the most interesting OSA subsets and tested for association only in this group with APL. There were 92 total families included in this analysis from the DBP low, BMI high, and IOP high subsets. Haplotype association tests for each gene were performed using a 2-SNP sliding window in APL.

Statistical Analyses in the Case-control Dataset

Single-marker association was determined by chi-square analysis of 2x2 (allelic) or 2x3 (genotypic) contingency tables. In addition, we tested association in 99 age-matched pairs with complete genotype data at all SNPs in the candidate genes and in subsets of the dataset stratified by CFH Y402H, LOC387715 A69S, a combination of CFH Y402H and LOC387715 A69S risk alleles, CFB R32Q, and BMI. We also used Haplo.stats to test 2-SNP sliding windows of haplotypes across each gene for association, and the haplo.glm module of haplo.stats to follow-up interesting haplotype results by weighted logistic regression, controlling for age, CFH Y402H, LOC387715 A69S, CFB R32Q, and smoking. Because phase was uncertain for some individuals, the posterior probability of a given haplotype for an individual served as a weight in the logistic regression.

All analyses were performed in grades 345vs12, but due to reduced sample size only two-point LOD scores and allelic and genotypic tests of association were carried out in the 5vs1 at least one eye and 5vs1 both eye subgroups.

Results

Linkage Disequilibrium

The linkage disequilibrium patterns across each gene for the case-control dataset grades 345vs12 are presented in Figures 6-2 to 6-5. Results in the cases and control separately, in the family-based dataset, and for the neovascular AMD analysis were similar (data not shown). In these figures, the darker the shading, the stronger the linkage disequilibrium among SNPs. The high amount of shading spanning nearly the full length of CACNG3, IL4R, and Q7Z6F8 suggests that we captured a substantial portion of the common variation in each gene. The LD between SNPs in HS3ST4 was weaker, as expected, since this gene is much larger than the others, covering more than 440 Kb.

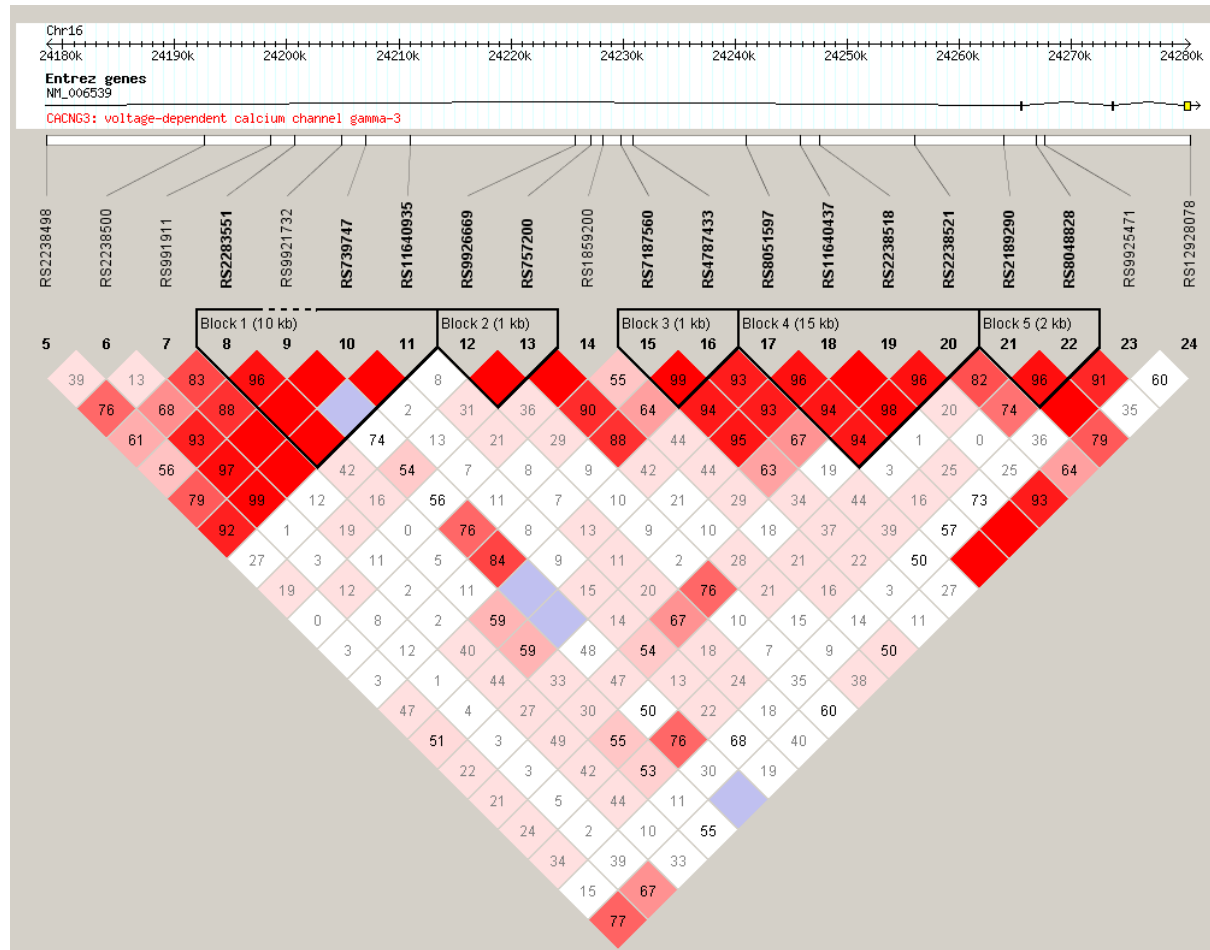


Figure 6-2. Linkage Disequilibrium in CACNG3 Case-control Grades 345vs12

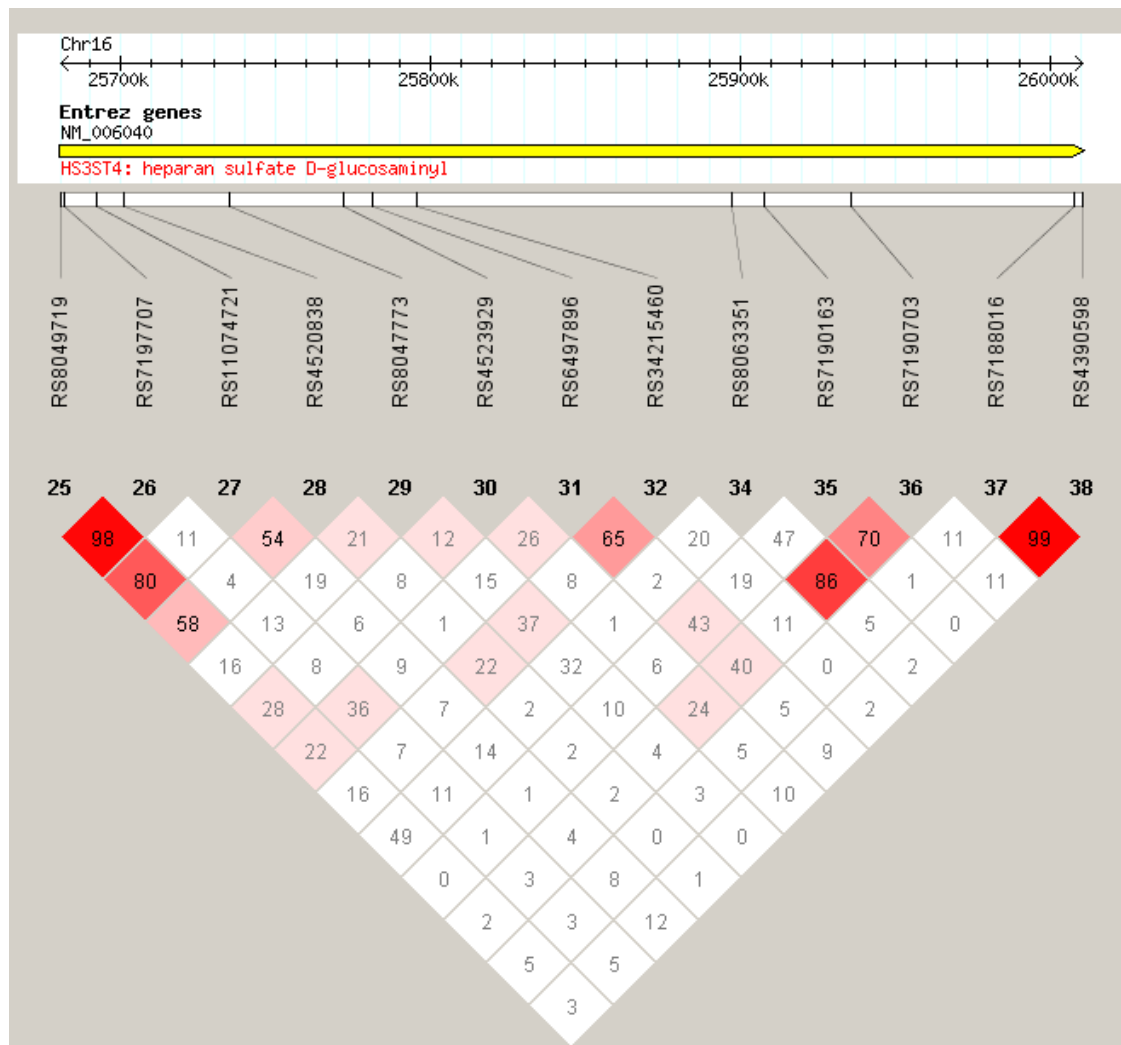


Figure 6-3. Linkage Disequilibrium in HS3ST4 Case-control Grades 345vs12

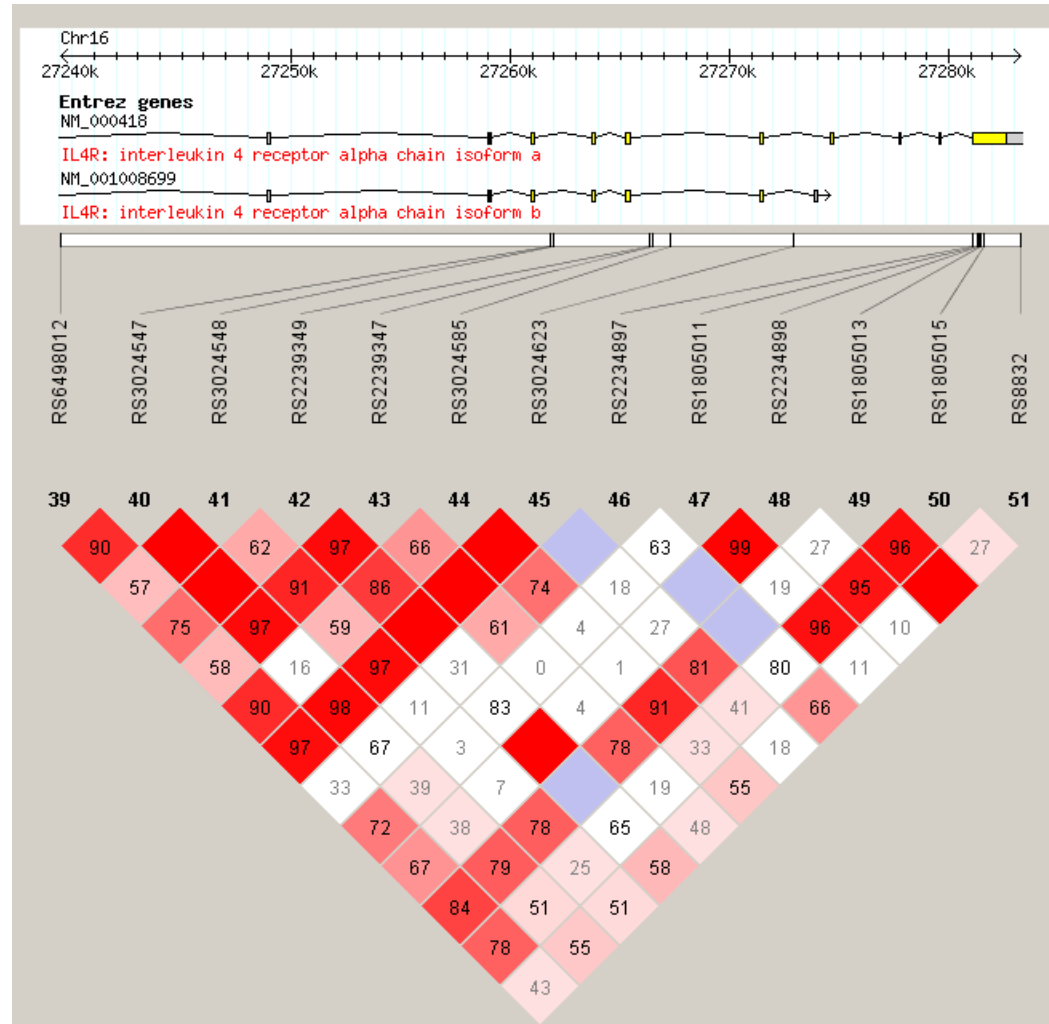


Figure 6-4. Linkage Disequilibrium in IL4R Case-control Grades 345vs12.

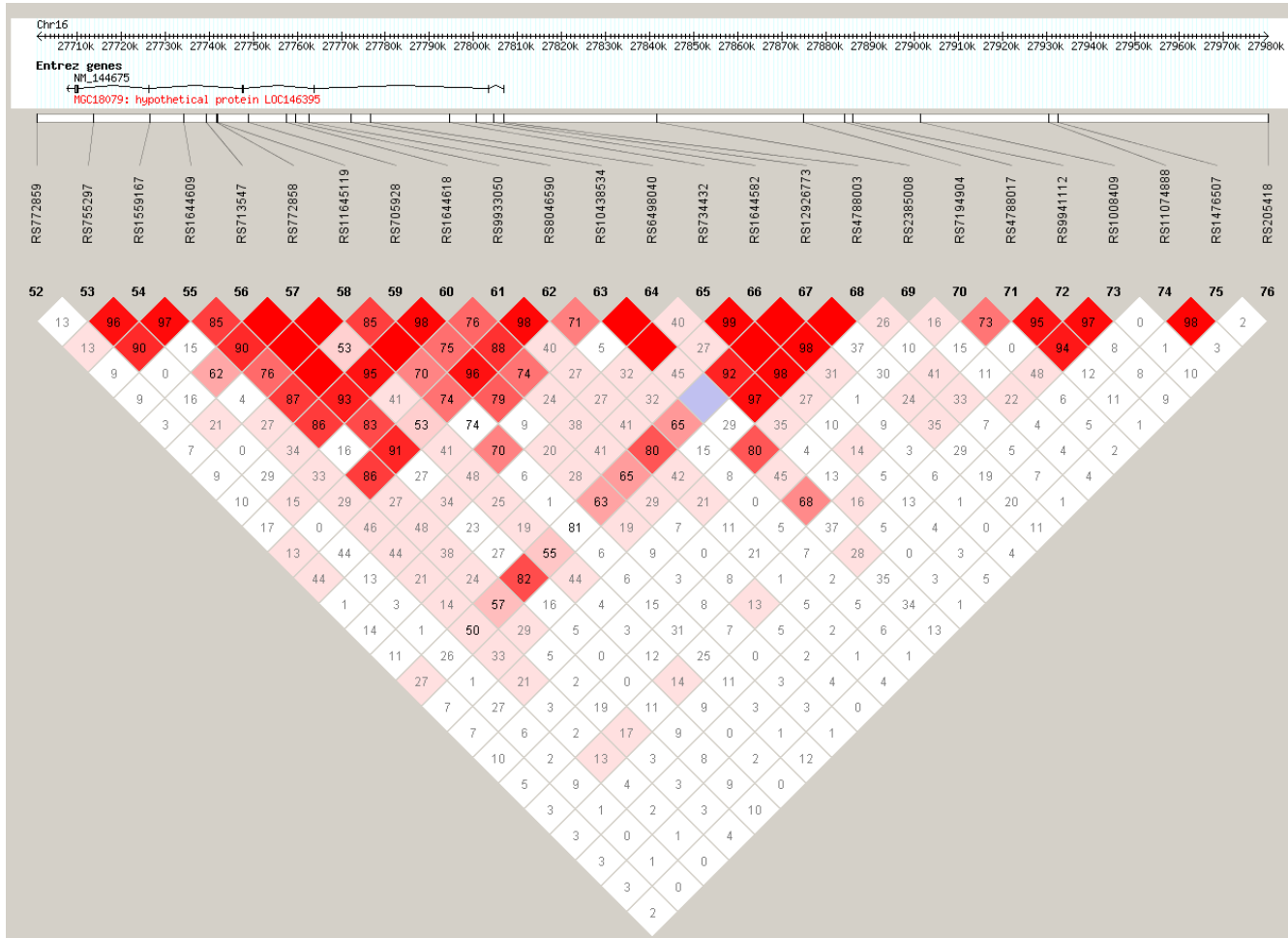


Figure 6-5. Linkage Disequilibrium in Q7Z6F8 Case-control Grades 345vs12.

Linkage

With additional multiplex families, the peak multipoint LOD score in the region decreased to 1.6 at 22.8 Mb under the dominant model (Figure 6-6). The nonparametric and recessive multipoint linkage curves followed the same shape as the dominant model, but were weaker in magnitude. However, CACNG3 showed even stronger 2pt linkage to AMD (peak 2pt NPL LOD = 3.34 at rs757200 grades 345vs12 compared to LOD=3.06 in screening analyses). This same SNP produced LOD scores near 1.0 in the recessive linkage analysis comparing grade 5 at least one eye vs grade 1, and another SNP in CACNG produced LOD scores greater than 2.0 in the grade 5 both eyes vs. grade 1 analysis. None of the other candidate genes yielded LOD scores greater than 2.0 in any analysis.

Because several genetic and environmental factors are either known or hypothesized to influence AMD susceptibility, we examined the evidence for linkage to chromosome 16 in various subsets of the families after ranking them from low to high or high to low on these factors. None of the covariates (CFH Y402H, LOC387715 A69S, CFB R32Q, smoking, blood pressure, IOP, and BMI) significantly increased the LOD score (Table 6-5). The greatest change was an increase of 1.30 LOD score units to a max LOD score of 1.54 at 26.0 Mb in the families with high IOP ($p=0.10$). Families with higher average BMI also appeared more strongly linked to this region, though there was not a significant difference in the LOD scores (max LOD 1.34 at 25.7 Mb, $p=0.19$). Encouragingly, these results show the same trend as previously reported by our group for this region (Schmidt et al. 2004), even though we have more than doubled the number of families in the sample from 62 to 131 multiplex families.

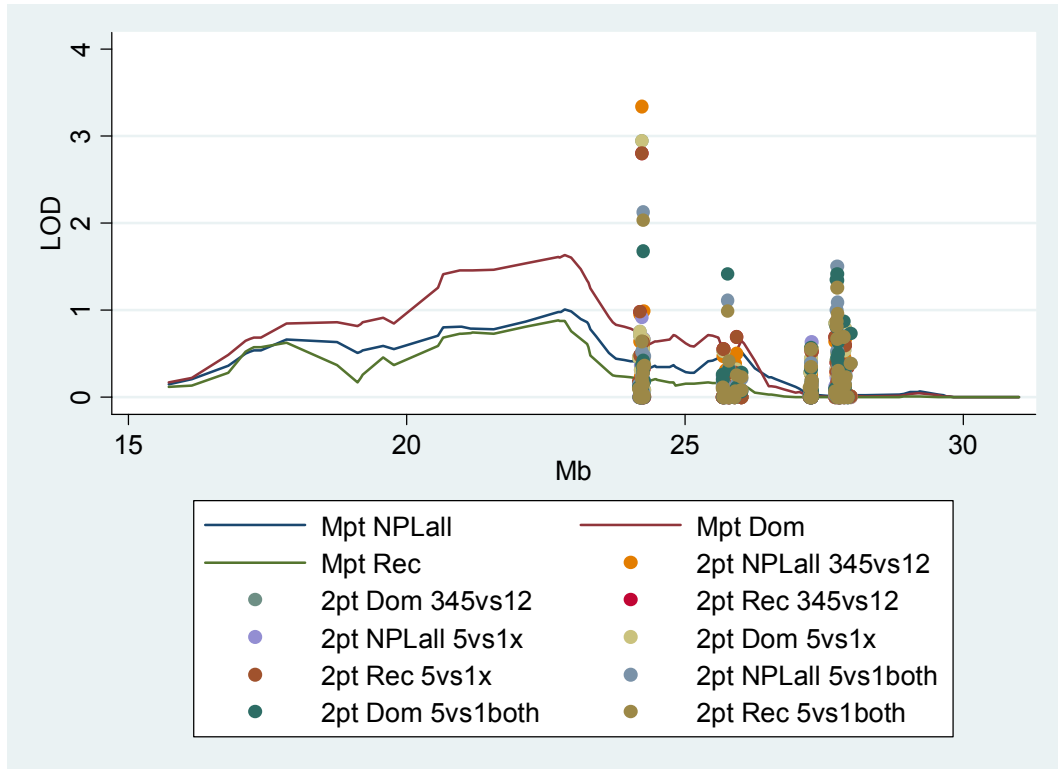


Figure 6-6. Linkage Analysis for the Candidate Genes on Chr. 16p12

Table 6-5. Ordered Subset Analysis on Chromosome 16p12. CFH Y402H and LOC387715 A69S covariates were weighted as the proportion of affected individuals in the family who carry the risk allele. Smoking was measured in pack-years and non-smokers were coded as having 0 pack-years. DBP=diastolic blood pressure, SBP=systolic blood pressure, IOP=intraocular pressure, BMI=body mass index, Ranking L=low to high, Ranking H=high to low. Note that even though we have 144 Mx families in the dataset, not all of them can be used in OSA due to missing covariate information.

| Variable | Ranking | Mb | Max LOD | Unconditional LOD | Delta LOD | Permutations | Empirical p-value | Families Used | Total Families | Proportion of Families |
|----------------|---------|------|---------|-------------------|-----------|--------------|-------------------|---------------|----------------|------------------------|
| CFH Y402H | L | 24.2 | 1.04 | 0.41 | 0.63 | 100 | 0.26 | 73 | 131 | 0.56 |
| CFH Y402H | H | 24.2 | 0.54 | 0.41 | 0.12 | 100 | 0.69 | 111 | 131 | 0.85 |
| LOC387715 A69S | L | 24.2 | 0.57 | 0.41 | 0.16 | 100 | 0.62 | 114 | 131 | 0.87 |
| LOC387715 A69S | H | 24.2 | 0.41 | 0.41 | 0.00 | 100 | 0.90 | 131 | 131 | 1.00 |
| CFB R32Q | L | 24.2 | 0.92 | 0.41 | 0.51 | 220 | 0.08 | 114 | 131 | 0.87 |
| CFB R32Q | H | 24.2 | 0.41 | 0.41 | 0.00 | 100 | 0.87 | 131 | 131 | 1.00 |
| Smoking | L | 24.2 | 0.80 | 0.41 | 0.39 | 100 | 0.44 | 114 | 131 | 0.87 |
| Smoking | H | 24.2 | 0.52 | 0.41 | 0.11 | 100 | 0.78 | 54 | 131 | 0.41 |
| DBP | L | 24.2 | 1.30 | 0.41 | 0.89 | 130 | 0.13 | 26 | 131 | 0.20 |
| DBP | H | 24.2 | 0.80 | 0.41 | 0.39 | 100 | 0.38 | 13 | 131 | 0.10 |
| SBP | L | 24.2 | 0.52 | 0.41 | 0.11 | 100 | 0.80 | 16 | 131 | 0.12 |
| SBP | H | 24.2 | 0.69 | 0.41 | 0.28 | 100 | 0.57 | 67 | 131 | 0.51 |
| IOP | L | 24.2 | 0.54 | 0.41 | 0.12 | 100 | 0.76 | 91 | 131 | 0.69 |
| IOP | H | 26.0 | 1.54 | 0.25 | 1.30 | 180 | 0.10 | 67 | 131 | 0.51 |
| BMI | L | 26.0 | 1.06 | 0.25 | 0.81 | 100 | 0.33 | 25 | 131 | 0.19 |
| BMI | H | 25.7 | 1.34 | 0.19 | 1.15 | 100 | 0.19 | 20 | 131 | 0.15 |

Association Analysis for CACNG3

SNPs in CACNG3 were most consistently associated with AMD (Figures 6-7, 6-8, and 6-9). Two-SNP haplotypes in both the family-based and case-control datasets were nominally statistically significant (APL rs4787433-rs8051597, case-control rs739747-rs11640935 and rs11640935-rs9926669, all $p < 0.01$, Figure 6-11). Consistent with an independent effect of this locus from CFH, 4 SNPs in CACNG3 were associated at the $p < 0.01$ level in either the non-CFH Y402H subset or the CFH Y402H/LOC387715 A69S combination low risk subset (Figure 6-10). This data is in agreement with OSA results from the original screening analyses in which linkage to chromosome 16 tended to be strongest in the subset of families with low rankings for CFH Y402H (LOD=2.56, $p=0.11$ Chapter V). Contrary to this hypothesis, rs1859200 was associated in the combination high subset ($p=0.003$). One SNP was also associated in the CFB R32Q protective allele carriers (rs11640935 $p=0.0001$). Examining the neovascular AMD cases, only genotype at rs7187560 was associated with AMD ($p=0.005$, Figure 7-8).

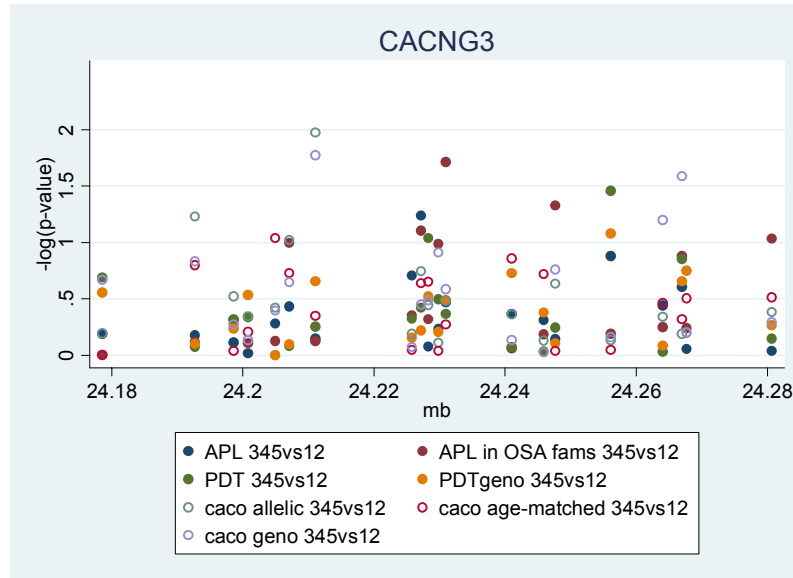
To more thoroughly investigate a potential haplotype effect in the CACNG3 gene, we selected the 3 SNPs that were nominally significant in the case-control sliding window haplotype analysis and tested those haplotypes for association. The ATA haplotype at SNPs rs739747, rs11640935, and rs9926669 was associated with decreased AMD risk, both in the complete case-control dataset and in the reduced sample with complete covariate data ($p=0.001$ for both). This haplotype was nearly twice as frequent in controls compared to cases (10.3% controls vs. 6.1% cases). After including the ATA haplotype in the context of known AMD susceptibility modifiers (age, CFH Y402H, LOC387715

A69S, CFB R32Q, and smoking), the haplotype was still strongly associated ($p=0.004$, $OR=0.38$, 95% confidence interval = 0.19 to 0.73).

Association Analysis for Q7Z6F8, HS3ST4, and IL4R

Three SNPs in Q7Z6F8 produced interesting results, though only in the case-control dataset and only when subsetting (rs755297 BMI<30 subset $p<0.01$, rs10438534 CFH Y402H risk allele carriers $p<0.01$, Figure 6-10, and rs734432 allelic association $p=0.006$ in 99 age-matched pairs, Figure 6-7). None of the other SNPs in Q7Z6F8 and no SNPs in IL4R or HS3ST4 produced a $LOD>2.0$ or association $p<0.01$, regardless of AMD subtype tested, age-matching, or stratification by CFH Y402H, LOC387715 A69S, CFB R32Q, smoking, or BMI.

A.



B.

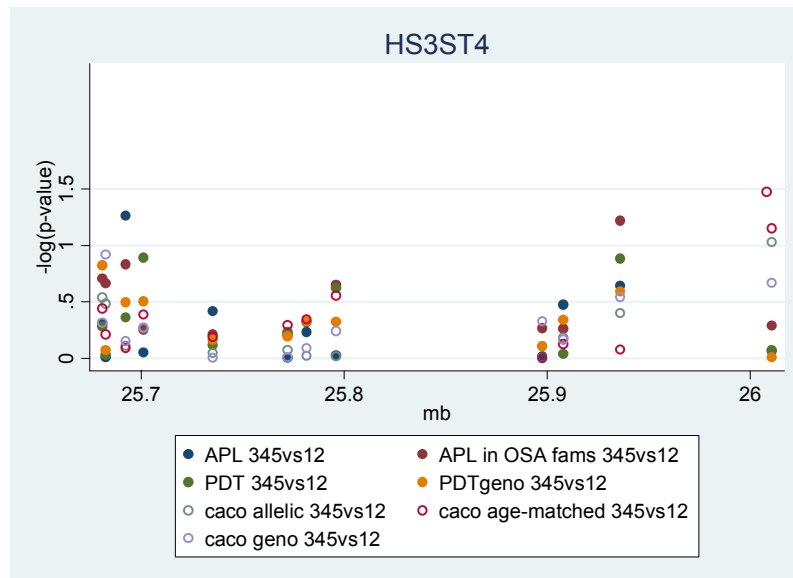
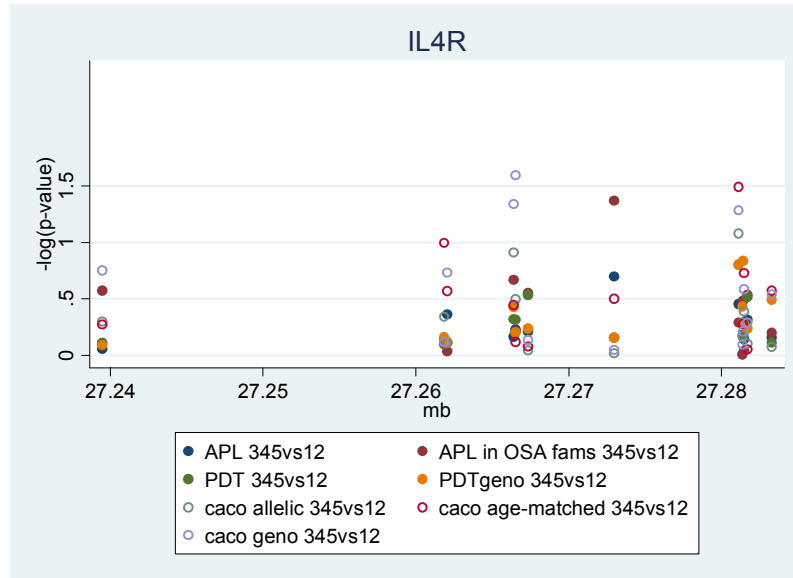


Figure 6-7. Family-based and Case-control Association Analysis in the Candidate Genes Grades 345vs12.

C.



D.

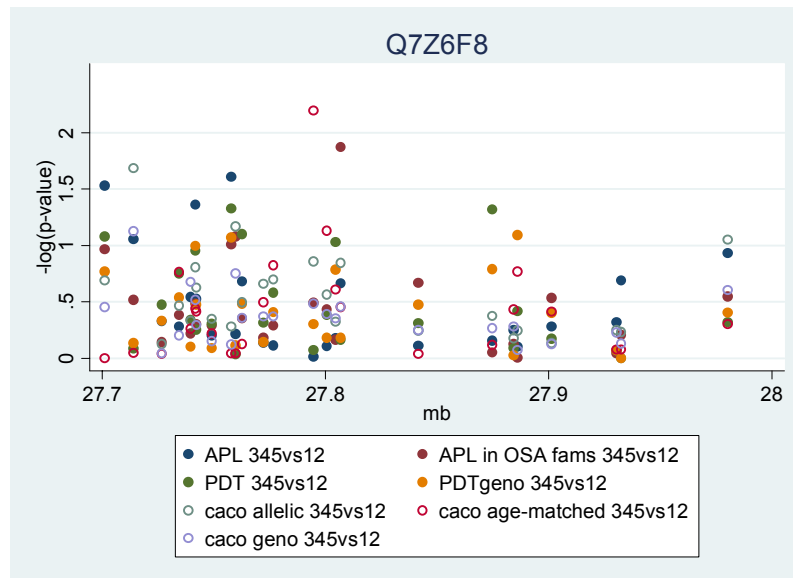
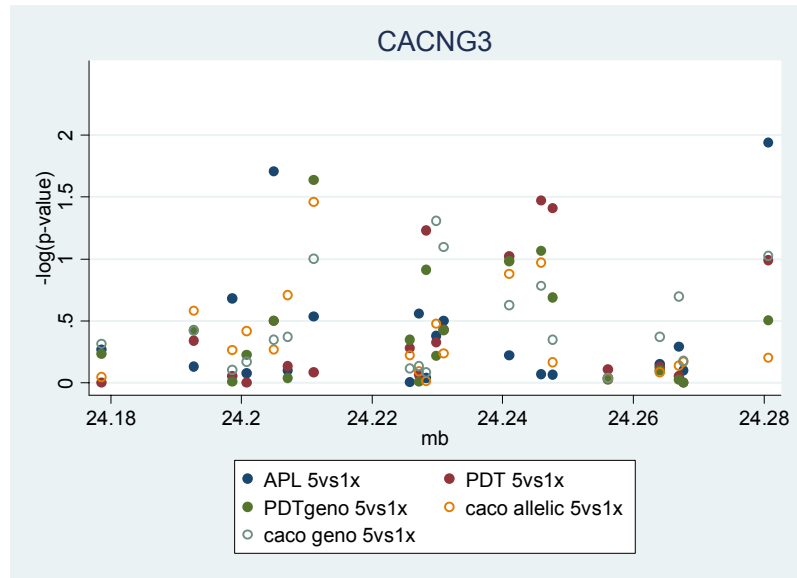


Figure 6-7cont.

A.



B.

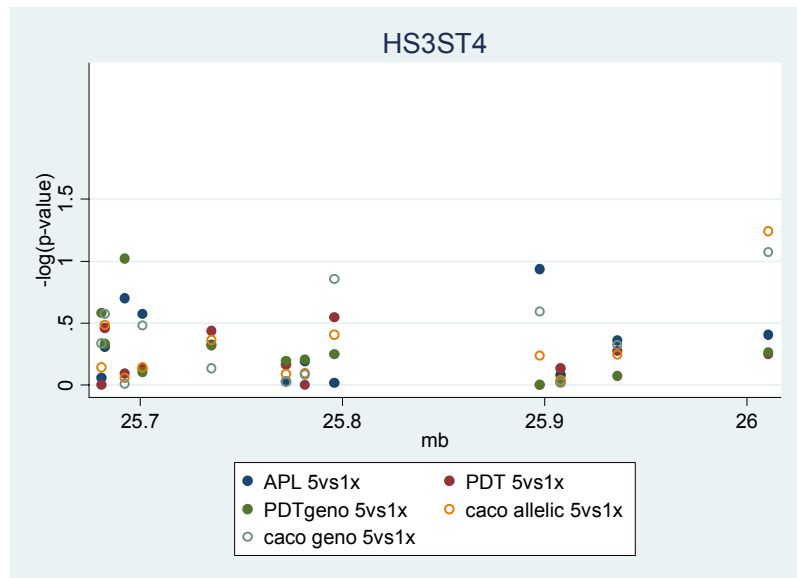
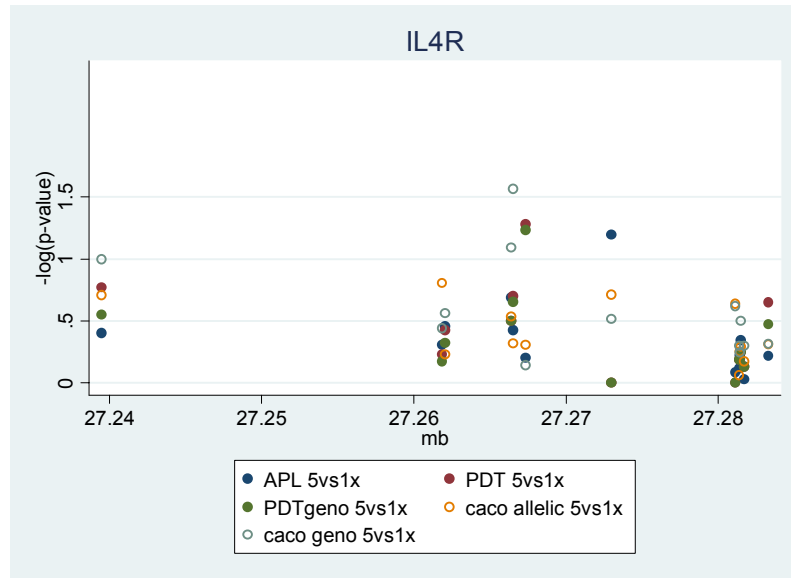


Figure 6-8. Family-based and Case-control Association Analysis in the Candidate Genes Grades 5vs1 at Least One Eye

C.



D.

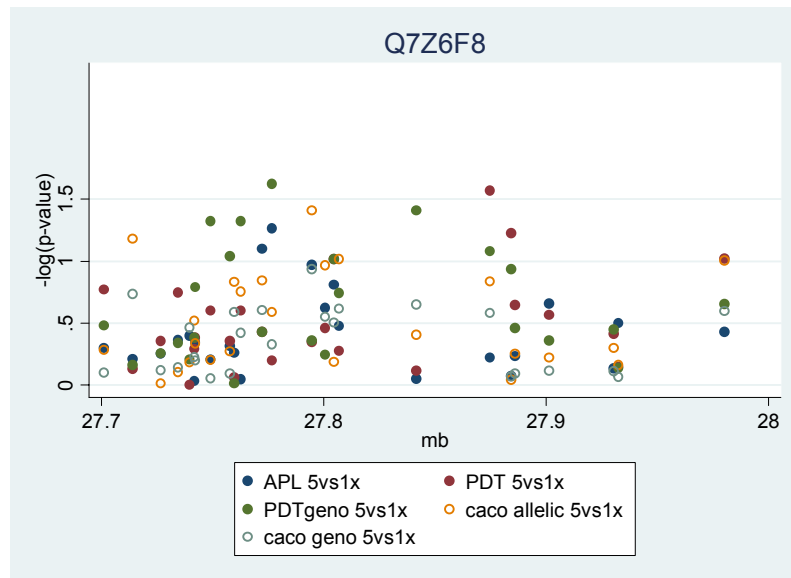
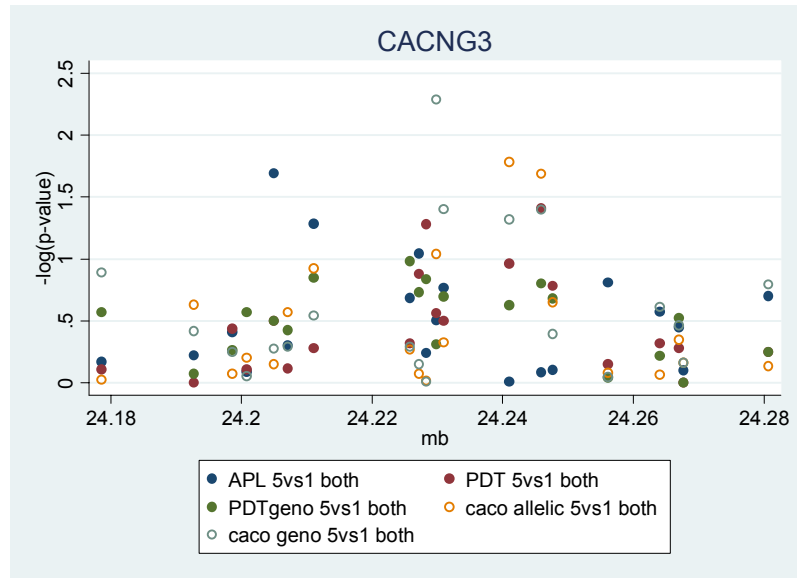


Figure 6-8 cont.

A.



B.

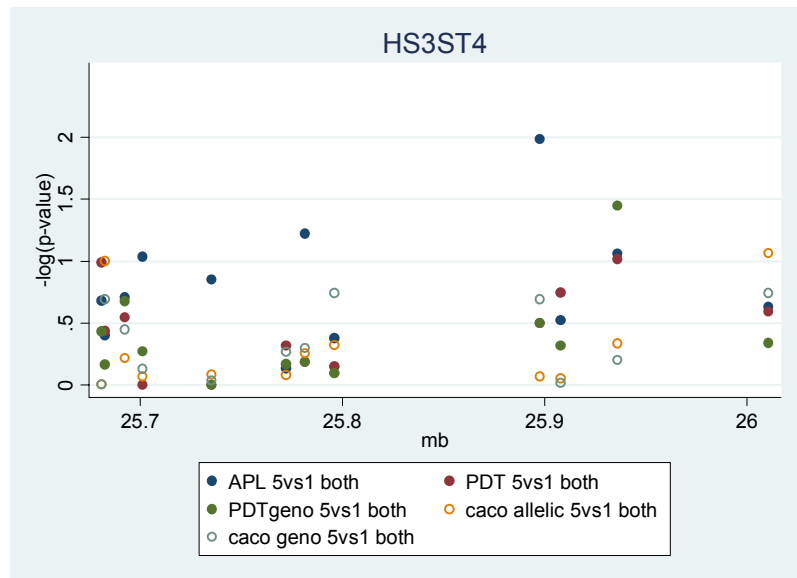
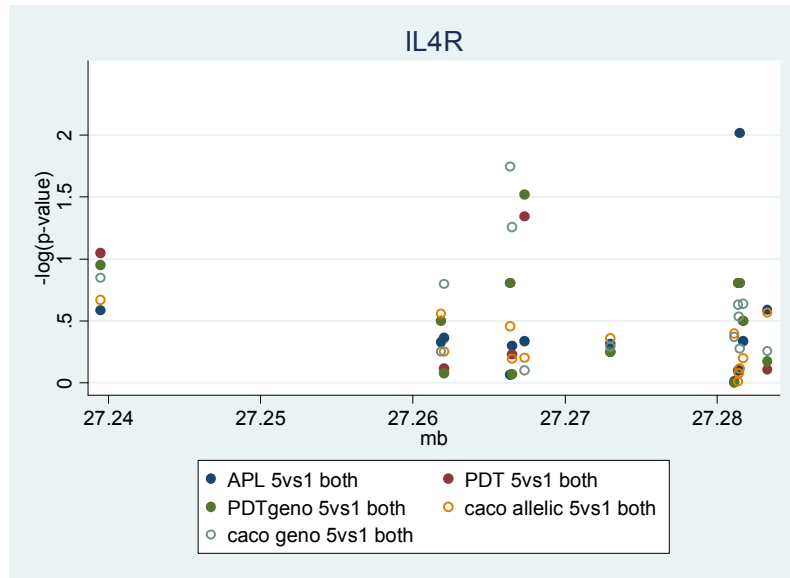


Figure 6-9. Family-based and Case-control Association in the Candidate Genes Grades 5vs1 Both Eyes

C.



D.

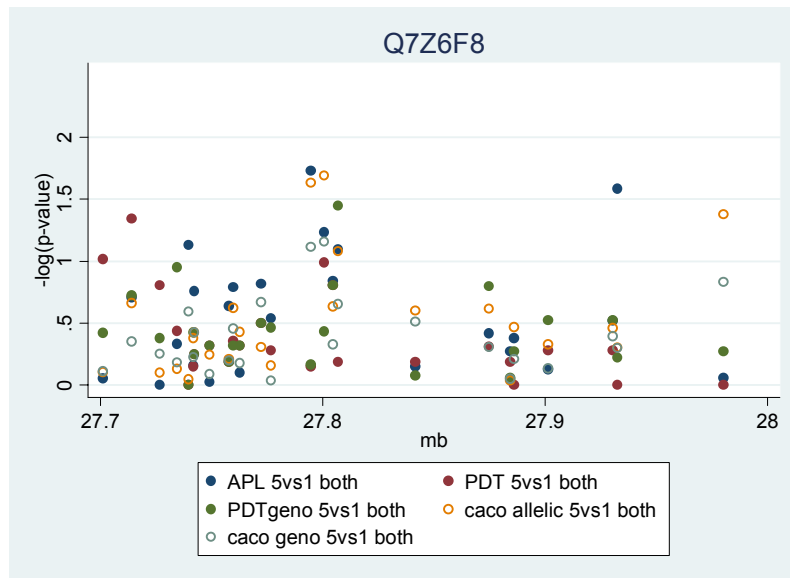
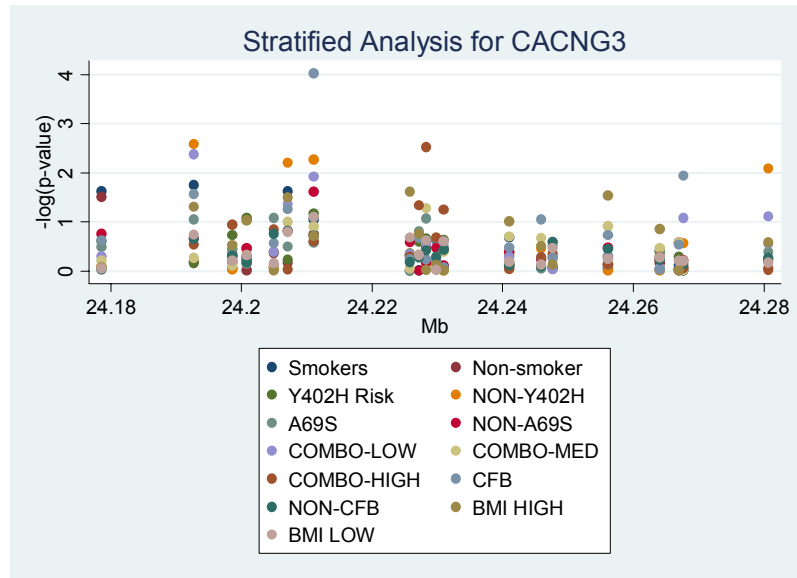


Figure 6-9 cont.

A.



B.

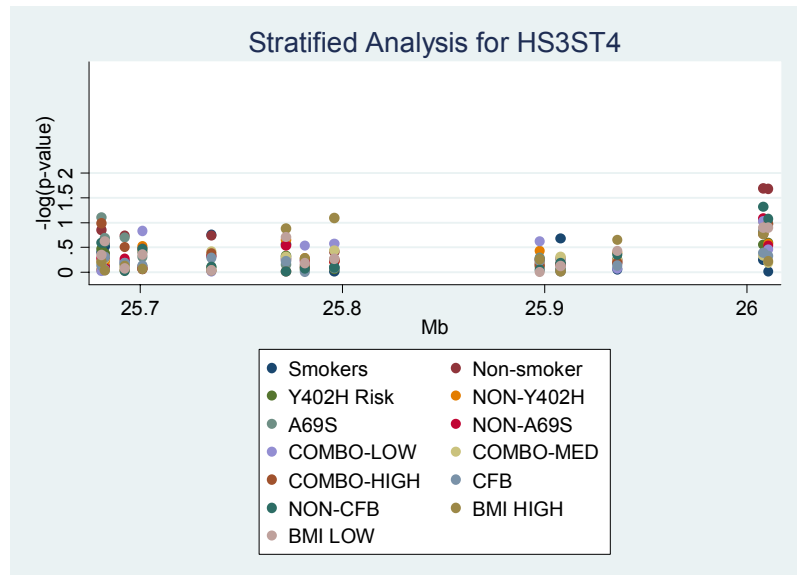
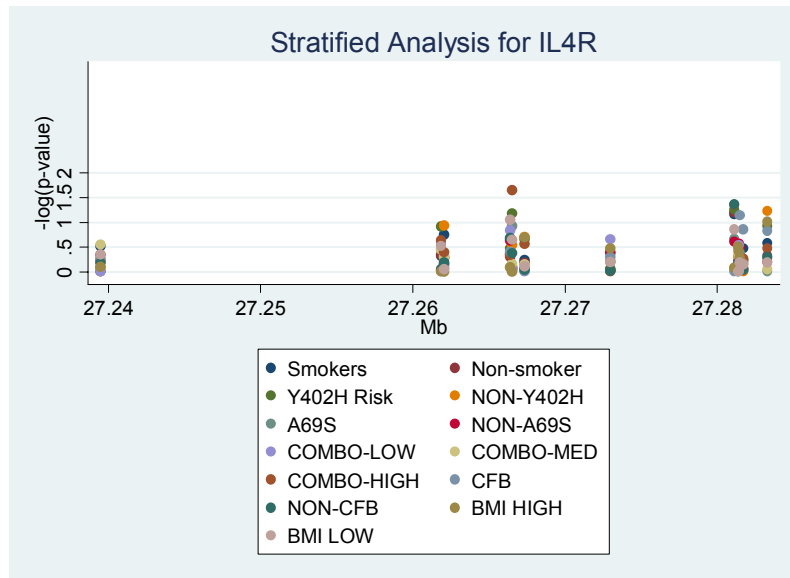


Figure 6-10. Stratified Case-control Association Analysis in the Candidate Genes Grades 345vs12

C.



D.

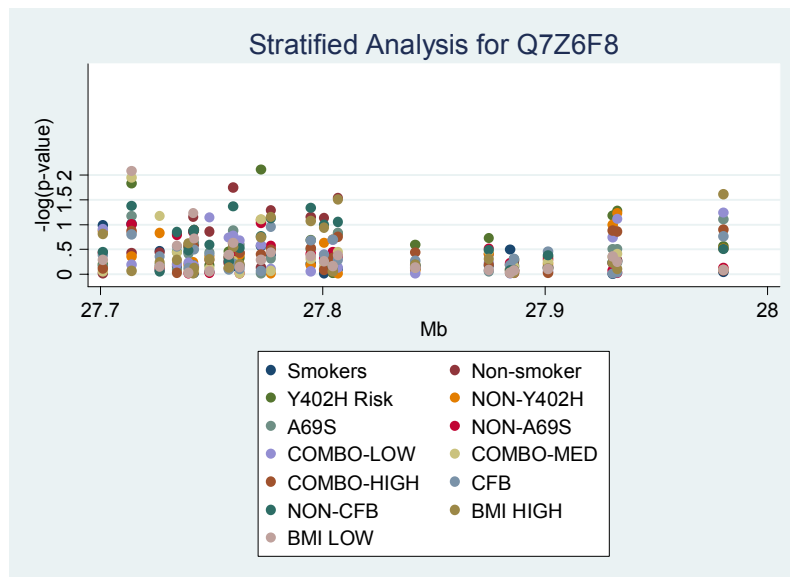


Figure 6-10 cont.

A.

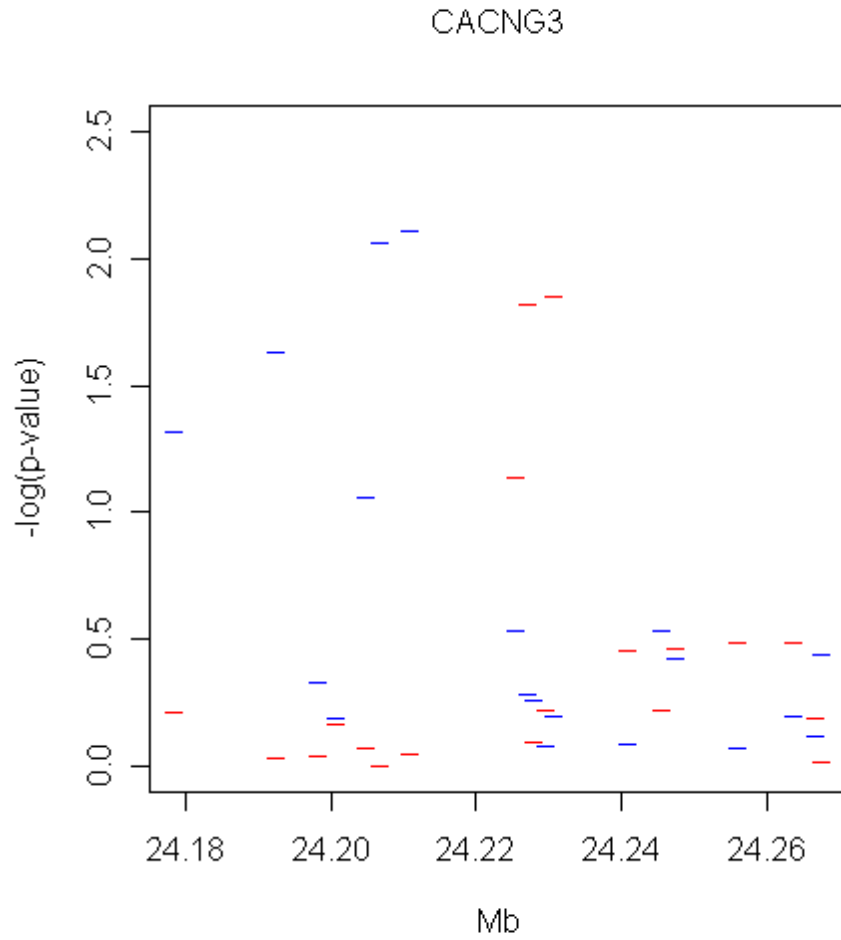


Figure 6-11. Family-based and Case-control Haplotype Analysis in the Candidate Genes Grades 345v12. Blue lines are for the case-control dataset, red lines are for the families.

B.

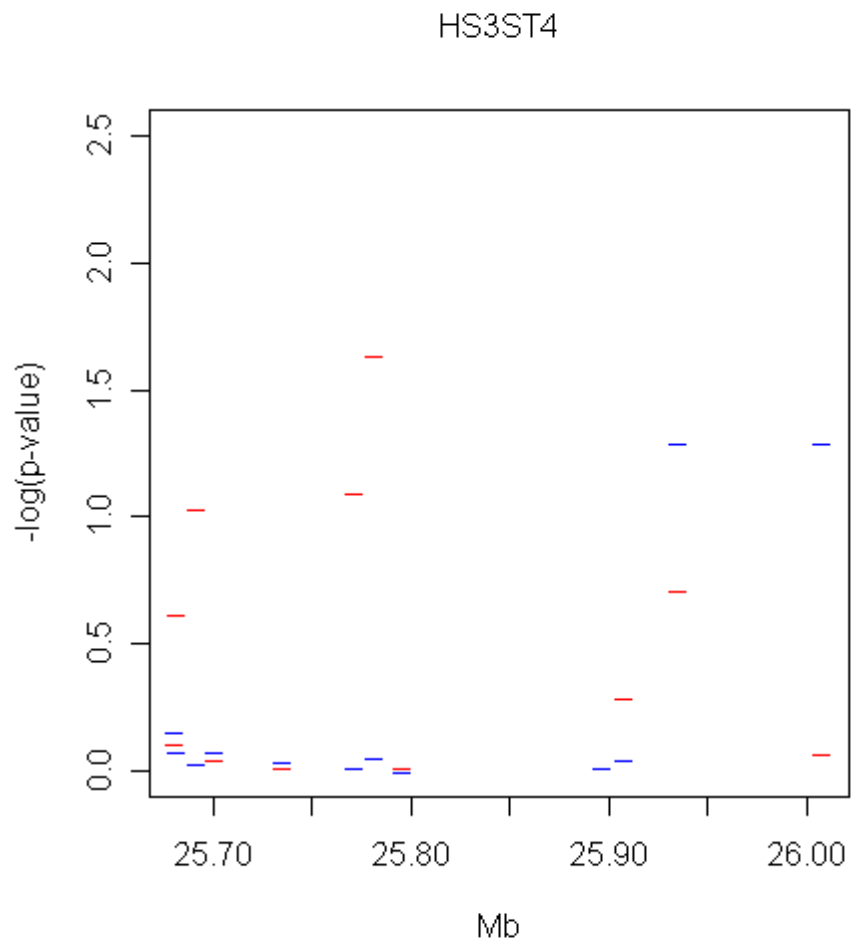


Figure 6-11 cont.

C.

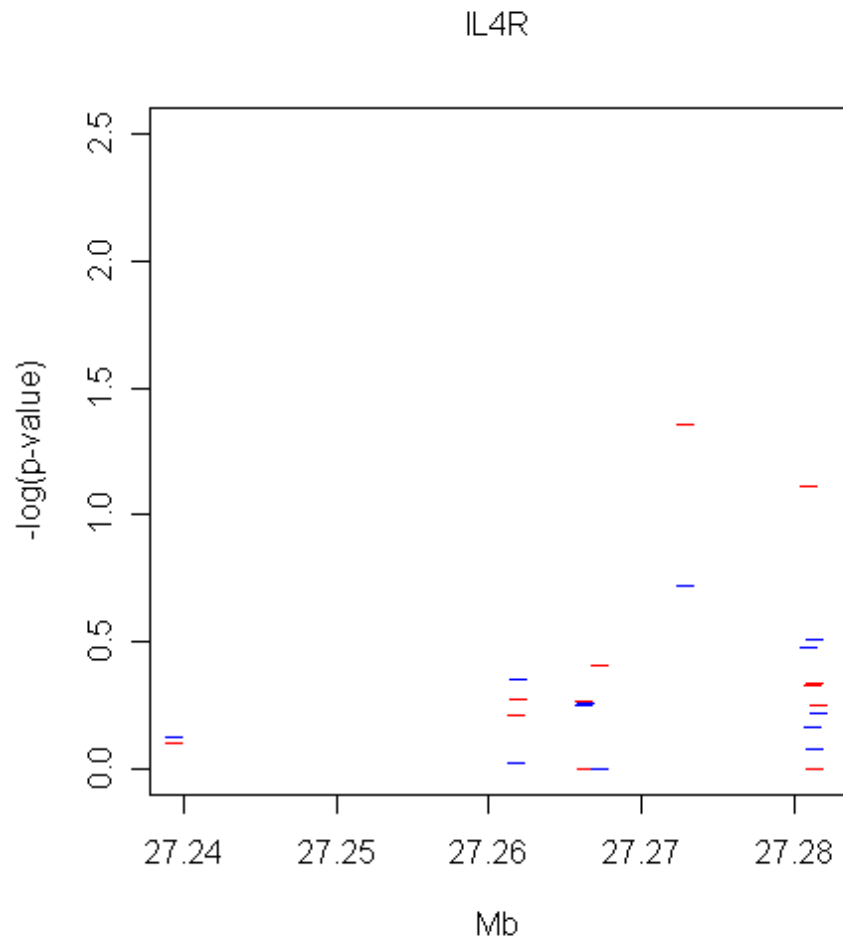


Figure 6-11 cont.

D.

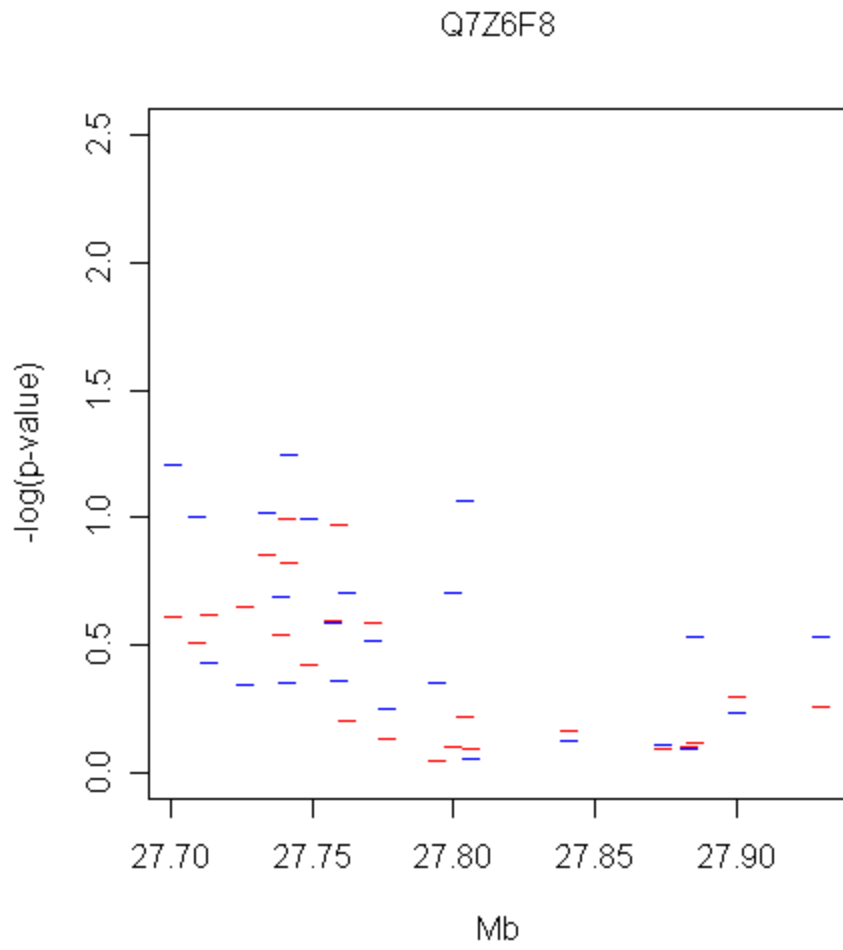


Figure 6-11 cont.

Discussion

SNPs in CACNG3 were strongly linked to AMD. Furthermore, the CACNG3 ATA haplotype was significantly more frequent in controls than cases, suggesting a protective effect against AMD or a risk effect of the major allele. Of the 3 SNPs composing the haplotype, only rs11640935 was associated with AMD by itself (allelic $p=0.01$, genotypic $p=0.02$). However, rs11640935 alone was not significant after including age, CFH Y402H, LOC387715 A69S, CFB R32Q, and smoking in the logistic model ($p=0.26$). Of the 4 SNPs showing association in CACNG3 in the CFH Y402H non-risk group or the CFH Y402H/LOC387715 A69S low combination, the minor alleles at 3 SNPs were more common in controls than cases, supporting the inverse association. These data argue that the association at CACNG3 is derived either from a true haplotypic effect or from a variant carried on this haplotype that has not yet been identified.

Intriguingly, a duplication spanning nearly 181 Kb and covering the 5' end and first exon of CACNG3 was recently discovered (Figure 6-12) (Wong et al. 2007). This duplication was present in 13 of 95 samples obtained from blood donors, the British Columbia Cancer Agency, and Coriell Cell Repository. The screening set was chosen to maximize ethnic diversity in the sample, and contains individuals from more than 17 distinct populations. The first SNP in the strongly associated ATA haplotype is only 29 Kb away from the proposed 3' end of the duplication. In the future, we plan to screen our AMD datasets for this duplication both to estimate its frequency in a Caucasian population and also to test it for association with AMD.

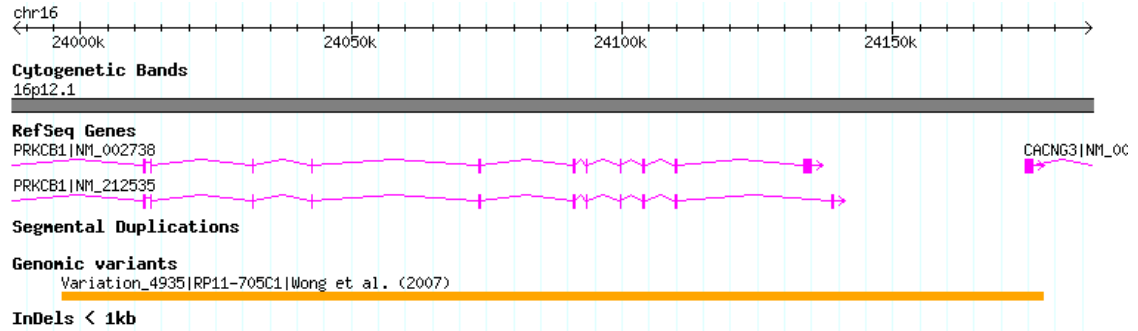


Figure 6-12. Duplication of PRKCB1 and CACNG3. This duplication covers the 5' upstream sequence and first exon of CACNG3, as well as the latter half of the PRKCB1 gene. PRKCB1 is expressed in the retina (NEI SAGE data) and involved in such diverse cellular processes as B cell activation, apoptosis induction, endothelial cell proliferation, and intestinal sugar absorption (NCBI Entrez Gene). One SNP in this gene was associated with neovascular AMD in the case-control dataset (rs916677 allelic $p=0.002$, genotypic $p=0.005$), but there were no interesting results for this gene in the family-based dataset. If this duplication is later associated with AMD, we will also consider PRKCB1 as a potential candidate gene.

Even though we did not see strong evidence of association for HS3ST4, IL4R, and Q7Z6F8, we cannot completely rule out these genes, especially since we estimated only moderate coverage of HS3ST4 and Q7Z6F8 given the current SNP density. We also did not screen these genes for multiple rare variants or copy number variants, which may be associated with AMD.

Interpretation of association results in light of multiple testing can be difficult, and statisticians disagree on the optimal way to correct for multiple comparisons. The most common method, the Bonferroni correction, and its modified form the step-down Bonferroni, both assume that each test is independent of all others. Clearly, this assumption is violated when SNPs in linkage disequilibrium are tested, leading to an overly conservative correction. Rather than apply too stringent a “correction” and miss

true positive results, we have chosen to report the nominal p-values, emphasizing that care be taken with their interpretation. We have called any LOD score >2.0 or p-value <0.01 “interesting” or “nominally significant”, but all of these associations should be considered tentative, pending replication in independent datasets.

In conclusion, CACNG3 is a candidate for the AMD locus on chromosome 16p12 based on its strong linkage and association results and a plausible biological function related to AMD. We plan to extensively follow-up these results by screening the AMD dataset for the duplication and genotyping additional SNPs near the 5' end of the gene.

CHAPTER VII

PREDICTIVE ACCURACY OF LOGISTIC REGRESSION AND IF-THEN DECISION TREE MODELS OF AMD

Introduction

The recent success in identifying both genetic and environmental modifiers of AMD susceptibility, which accounts for the largest number of cases of vision loss in the elderly, has prompted quantification of the percentage of AMD cases that can be explained by known risk factors. CFH Y402H, LOC387715 A69S, and cigarette smoking are now well-accepted risk factors for AMD, and CFB R32Q is associated with decreased AMD risk (Chapter I).

We have estimated that variation in CFH, LOC387715, and cigarette smoking together explain approximately 61% of the population- attributable risk (PAR) for AMD (Schmidt et al. 2006). This agrees well with the PAR of 63% estimated by Schaumberg et al., though they did not include smoking in their model (Schaumberg et al. 2007). Furthermore, SNPs in CFH and CFB combined result in a sibling recurrence risk (λ_s) of ~ 2.0 , a substantial portion of the overall AMD λ_s of $\sim 3-6$ (Maller et al. 2006).

Given these data, it is reasonable to ask whether knowledge of an individual's risk factor data can be used to make a prediction about his or her chance of developing AMD. Using genotype information at CFH, LOC387715, and CFB/CC2, AMD status was predicted correctly 74% of the time in cases and 56% of the time in controls (Gold et al. 2006). However, the predictive accuracy of a single model containing all well-replicated

AMD susceptibility factors (age, smoking, CFH Y402H, LOC387715 A69S, and CFB R32Q) has yet to be tested.

We chose to model these factors in both logistic regression and an if-then decision trees in two distinct datasets. First, we constructed these models in half the Vanderbilt/Duke case-control study population and then tested the models on the other half of this dataset. For a more rigorous test of how the model would apply to population-based data we then built the model in the complete Vanderbilt/Duke population and applied it to the prospective AMD cohort in the Memphis Health ABC Study. Sensitivity, specificity, and overall correct classification rate were used to evaluate the success of the models.

Materials and Methods

Study Populations

It is essential to construct a predictive model in one dataset and then apply the model in a separate dataset to avoid bias in model evaluation. Additionally, large variation in the demographic characteristics of the training and testing datasets may reduce the number of individuals correctly classified in the testing dataset. Therefore, we carefully compared the testing and training datasets to determine if they differed substantially for any important demographic traits (Tables 7-1 and 7-2). It is important to note that the Vanderbilt/Duke dataset was ascertained as a clinic-based AMD population, whereas the Memphis samples were part of a prospective cohort group of over 3,000 individuals aged 70-79 years randomly selected from Medicare rolls in Memphis, TN or Pittsburgh, PA

who did not have difficulty walking a quarter of a mile or climbing a flight of stairs at the time of study enrollment (Iannaccone et al. 2007; Gallaher et al. 2007). The African American samples were removed from the Memphis analysis since there were no African Americans in the Vanderbilt/Duke dataset.

Table 7-1. Comparison of the Two Halves of the Vanderbilt/Duke Study Population

| Characteristic | Training | Testing |
|---|-----------------|----------------|
| Cases (#) | 165 | 168 |
| Controls (#) | 93 | 91 |
| Ascertainment Center (% Duke) | 56.2 | 57.5 |
| Age (mean) | 72.5 | 72.5 |
| Gender (% Female) | 62.4 | 57.1 |
| CFH Y402H (% Risk Allele Carriers) | 73.3 | 80.3 |
| LOC387715 A69S (% Risk Allele Carriers) | 45.3 | 45.6 |
| CFB R32Q (% Protective Allele Carriers) | 13.6 | 9.3 |
| Smokers (%) | 41.1 | 43.6 |

Table 7-2. Comparison of the Vanderbilt/Duke and Memphis Study Populations

| Characteristic | Vanderbilt/Duke (Training) | Memphis (Testing) |
|---|---------------------------------------|------------------------------|
| Cases (#) | 430 | 111 |
| Controls (#) | 222 | 214 |
| Age (mean) | 72.6 | 79.2 |
| Gender (% Female) | 59.2 | 51.4 |
| Race (% African American) | 0.0 | 19.7 |
| CFH Y402H (% Risk Allele Carriers) | 77.5 | 60.3 |
| LOC387715 A69S (% Risk Allele Carriers) | 55.2 | 42.2 |
| CFB R32Q (% Protective Allele Carriers) | 11.7 | 24.0 |
| Smokers (%) | 57.5 | 48.0 |

Building the Models

Logistic regression and if-then decision tree rules were used to build models of AMD

(Figure 7-1).

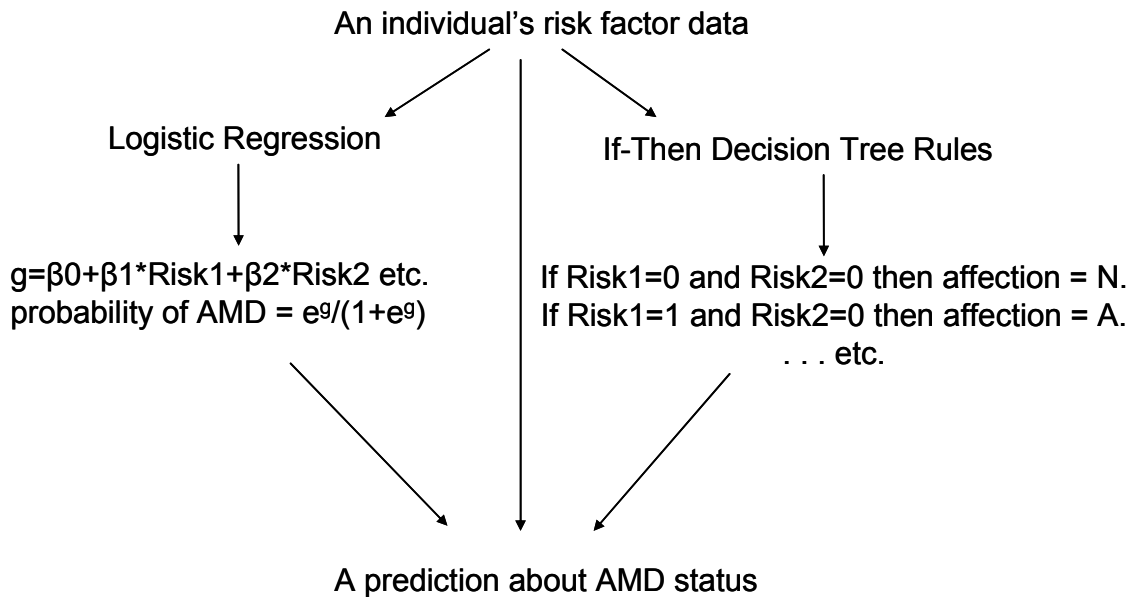


Figure 7-1. Two Approaches for Building the AMD Models.

For the logistic regression analyses, we included age (in years), ever/never smoking (coded “1” for smokers, “0” for non-smokers), CFH Y402H (coded “1” for CC and CT genotypes, “0” for TT genotype), LOC387715 A69S (coded “1” for TT and GT genotypes, “0” for GG genotype) and CFB R32Q (coded “1” for AA and AG genotypes and “0” for GG genotype) in the model. Therefore, the logistic regression equation was:

$$g = \beta_0 + \beta_1 * \text{Age} + \beta_2 * \text{Y402H} + \beta_3 * \text{A69S} + \beta_4 * \text{CFB} + \beta_5 * \text{Smoking}$$

and the probability of AMD for an individual was:

$$\text{probability of AMD} = \frac{e^g}{1 + e^g}$$

We did not include interactions terms in the model so that excess stratification of the dataset could be avoided.

Once the probability of AMD was determined for each individual in the testing dataset, individuals with a probability greater than a particular threshold were classified as affected, and those below the threshold were classified as normal. These “model calls” can then be compared to the affection status assigned by a clinician and the sensitivity, specificity, and overall correct classification rate of the model can be determined.

Changing the threshold for the probability of AMD will change the number of false positives and false negatives called by the model. Because there is no *a priori* reason to select a particular threshold value, we chose to use 0.5 as a cut-off for our analyses. After examining the histogram of AMD probabilities by true affection status, we raised the threshold to 0.75 in the Memphis testing dataset in an attempt to increase accuracy. Finally, we used ROC curves (plots of sensitivity vs. 1-specificity) to determine the threshold which would have correctly classified the greatest number of individuals.

For the if-then decision tree rules, the number of cases and controls with each particular susceptibility factor combination was calculated. If the ratio of cases to controls having this combination in the training dataset exceeded the total ratio of cases and controls, then individuals with the same combination in the testing dataset were called affected and vice versa. For example, there were 165 cases and 93 controls in the Vanderbilt/Duke training dataset for a total ratio of 1.77. There were 12 cases and 7 controls that were non-smokers and had CC, GG, and GG genotypes at CFH Y402H, LOC387715 A69S, and CFB R32Q, respectively, for a ratio of 1.71. Therefore, all individuals with this combination in the testing dataset were classified as controls.

One advantage of this type of modeling compared to logistic regression is that there is no need to specify an arbitrary threshold value for classifying affection status. The major drawbacks with if-then decision tree rules are: 1) age cannot be included in the model without overly stratifying the datasets and 2) large sample sizes are needed for each susceptibility combination to ensure stability of the model.

Evaluating the Models

Sensitivity, specificity, and overall correct classification rate were calculated according to the following equations:

$$\text{Sensitivity} = \text{Probability}(\text{Affected}_{\text{model}}|\text{Affected}_{\text{reality}})$$

$$\text{Specificity} = \text{Probability}(\text{Normal}_{\text{model}}|\text{Normal}_{\text{reality}})$$

$$\text{Overall Classification Rate} = (\# \text{ Cases Correct} + \# \text{ Controls Correct}) / \\ (\text{Total} \# \text{ of Individuals in the Testing Dataset})$$

Results

Vanderbilt/Duke Study Population

We first constructed a logistic model of AMD in half the Vanderbilt/Duke study population (Table 7-3) and then applied it to the other half of the dataset. Several factors that are usually strongly associated with AMD were not impressive in this analysis (CFH Y402H, LOC387715 A69S, and smoking), but this is probably caused by the small sample size of the training dataset.

Table 7-3. Logistic Regression Model in Half the Vanderbilt/Duke Training Dataset

| Parameter | Estimate | Std. Error | χ^2 | p-value | Odds Ratio | 95% Confidence Interval | |
|------------------|-----------------|-------------------|----------------------------|----------------|-------------------|--------------------------------|------|
| Intercept | -9.52 | 1.45 | 42.93 | <.0001 | | | |
| Age | 0.13 | 0.02 | 44.78 | <.0001 | 1.14 | 1.10 | 1.19 |
| CFH Y402H | 0.04 | 0.17 | 0.07 | 0.79 | 1.09 | 0.56 | 2.13 |
| LOC387715 A69S | 0.24 | 0.15 | 2.41 | 0.12 | 1.61 | 0.88 | 2.94 |
| CFB R32Q | -0.72 | 0.23 | 10.11 | <.0001 | 0.24 | 0.10 | 0.58 |
| Smoking | 0.21 | 0.16 | 1.77 | 0.18 | 1.52 | 0.82 | 2.80 |

Using a probability threshold of 0.5, this model of AMD was accurate 82% of the time (Table 7-4). Plotting the probabilities by true affection status revealed that most cases had probabilities closer to 1 and most controls had probabilities closer to 0, as expected (Figure 7-2).

Table 7-4. Evaluation of the Logistic Regression Model in the Vanderbilt/Duke Testing Dataset.

| Sensitivity | Specificity | Overall Correctly Classified |
|------------------|----------------|------------------------------|
| 145/168 (86%) | 66/91 (73%) | 211/259 (82%) |

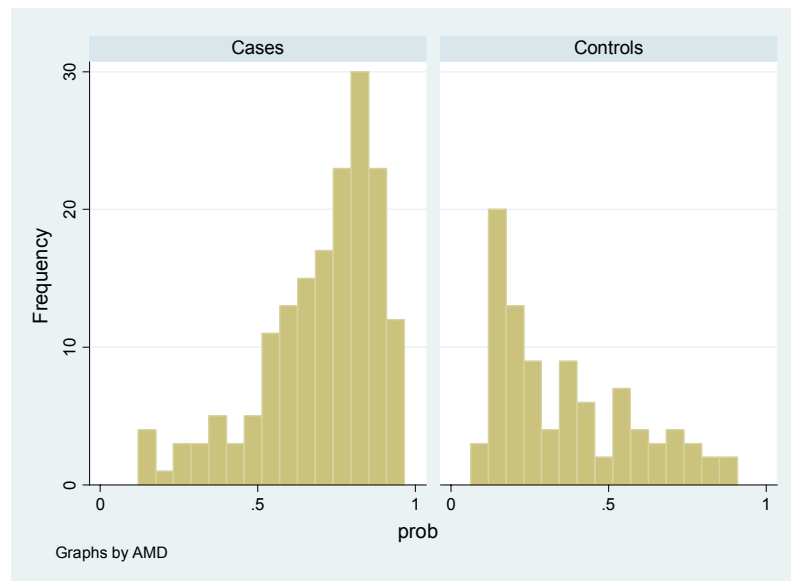


Figure 7-2. Histogram of Probabilities of AMD by Affection Status in the Duke/Vanderbilt Testing Dataset

According to the ROC curve, a threshold of 0.48 would have provided the maximum possible classification rate of 82.6% with 88.1% sensitivity and 72.5% specificity (Figure 7-3). The area under the ROC curve was 0.86 (95% confidence interval 0.81 to 0.91), suggesting that this model is significantly better than chance at predicting AMD status.

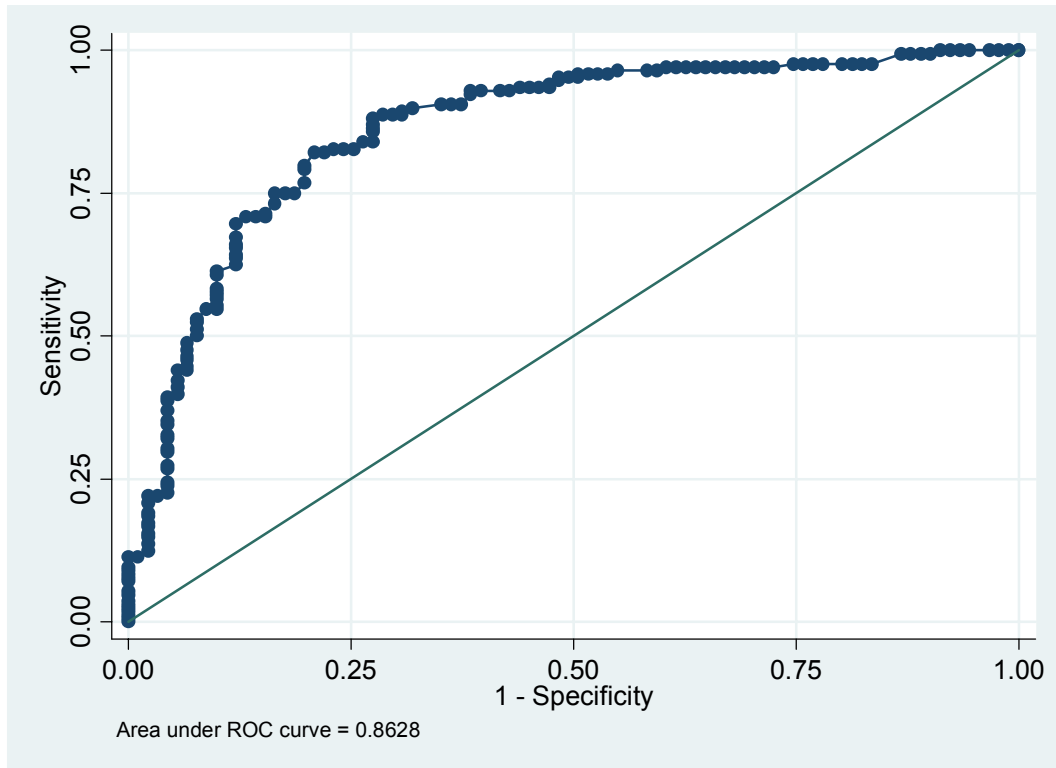


Figure 7-3. ROC Curve in the Vanderbilt/Duke Testing Dataset

Next, we used if-then decision tree rules to build a model of AMD. Several combinations of susceptibility factors were not observed in the training dataset (Figure 7-4). However, no one in the testing dataset carried these combinations, so the overall classification rate of the model was unaffected. The ratio of cases to controls in each cell

had to exceed 1.77 (the ratio of cases to controls in the training dataset) for a combination to be classified as affected. The if-then rules are depicted in Table 7-5.

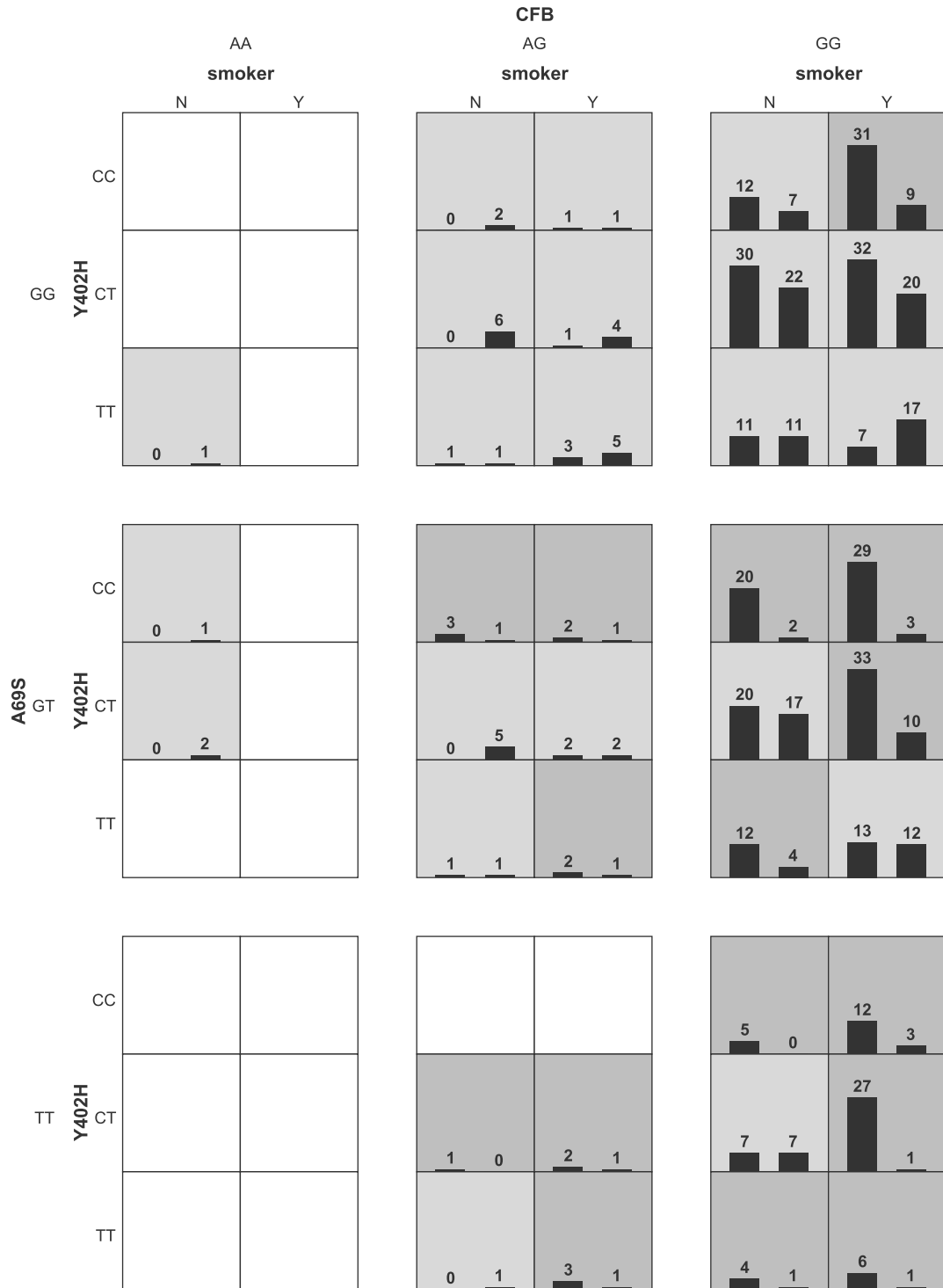


Figure 7-4. Counts of Cases and Controls for Each Susceptibility Combination in the Vanderbilt/Duke Training Dataset. Cases are plotted on the left, controls on the right of each square. Combinations leading to a model call of "affected" are shaded dark gray, and those resulting in a model call of "normal" are shaded light gray. White cells occurred when no one in the training dataset had that particular combination of susceptibility factors.

Table 7-5 If-Then Decision Tree Rules Developed in Half the Vanderbilt/Duke Training Dataset.

| |
|--|
| IF smoker = N AND Y402H = CC AND R32Q = AA AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CC AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CC AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = N AND Y402H = CC AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CC AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = N AND Y402H = CC AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = N AND Y402H = CT AND R32Q = AA AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = AA AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CC AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = CC AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Normal . |
| IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |
| IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected. |

The overall correct classification rate for the if-then decision tree rules was 63%, much lower than for the logistic regression (Table 7-6). This may be caused by instability of the decision tree model resulting from small numbers of observations for each susceptibility combination.

Table 7-6. Evaluation of the If-Then Decision Tree Rules in the Vanderbilt/Duke Testing Dataset.

| Sensitivity | Specificity | Overall Correctly Classified |
|--------------------|--------------------|-------------------------------------|
| 92/168 (55%) | 72/91 (79%) | 164/259 (63%) |

Memphis Study Population

Next, we built the logistic regression model in the Vanderbilt/Duke dataset and applied it to the Memphis dataset. Reassuringly, all the variables in the model were highly significant in the complete Vanderbilt/Duke Training Dataset (Table 7-7).

Table 7-7. Logistic Regression Model in the Full Vanderbilt/Duke Training Dataset

| Parameter | Estimate | Std. Error | χ^2 | p-value | Odds Ratio | 95% Confidence Interval | |
|------------------|-----------------|-------------------|----------------------------|----------------|-------------------|--------------------------------|------|
| Intercept | -10.78 | 0.97 | 122.87 | <.0001 | | | |
| Age | 0.15 | 0.01 | 128.60 | <.0001 | 1.16 | 1.13 | 1.19 |
| CFH Y402H | 0.38 | 0.12 | 10.62 | 0.001 | 2.14 | 1.36 | 3.39 |
| LOC387715 A69S | 0.39 | 0.10 | 14.42 | 0.0001 | 2.17 | 1.45 | 3.24 |
| CFB R32Q | -0.67 | 0.15 | 19.38 | <.0001 | 0.26 | 0.14 | 0.47 |
| Smoking | 0.27 | 0.10 | 6.82 | 0.009 | 1.71 | 1.14 | 2.56 |

Unfortunately, the logistic model did not accurately predict affection status regardless of the probability threshold used (Table 7-4). The probabilities of AMD did not separate cleanly into two distinct groups (Figure 7-5). Using the ROC curve to determine the best possible threshold, we could have correctly classified 67% of the Memphis samples with 31% sensitivity and 89% specificity, had we used a threshold of 0.895 (Figure 7-6). The area under the ROC curve was only 0.62 (95% confidence interval 0.54 to 0.69), confirming the poor predictive ability of this model in the Memphis population.

Table 7-8. Evaluation of the Logistic Regression Model in the Memphis Testing Dataset.

| Threshold | Sensitivity | Specificity | Overall Correctly Classified |
|-----------|----------------|-----------------|------------------------------|
| 0.50 | 84/85 (99%) | 2/148 (1%) | 86/233 (37%) |
| 0.75 | 67/85 (79%) | 42/148 (28%) | 109/233 (47%) |

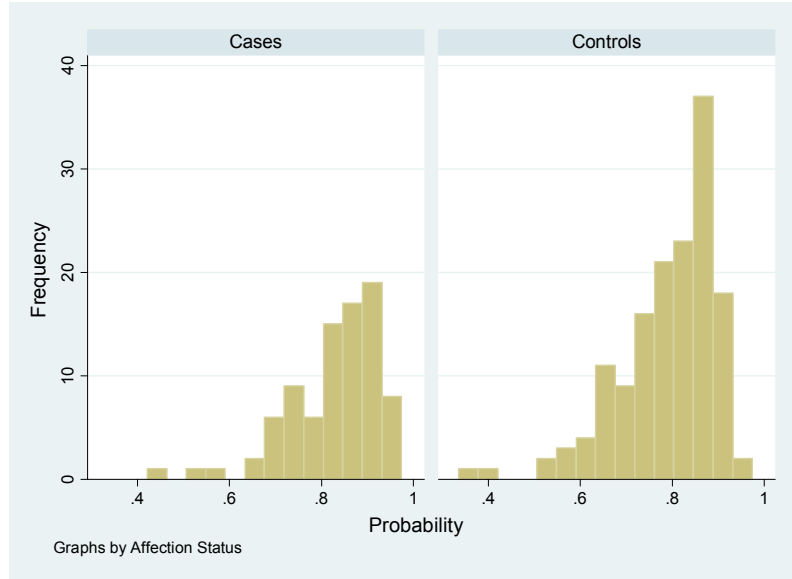


Figure 7-5. Histogram of Probabilities of AMD by Affection Status in the Complete Vanderbilt/Duke Training Dataset

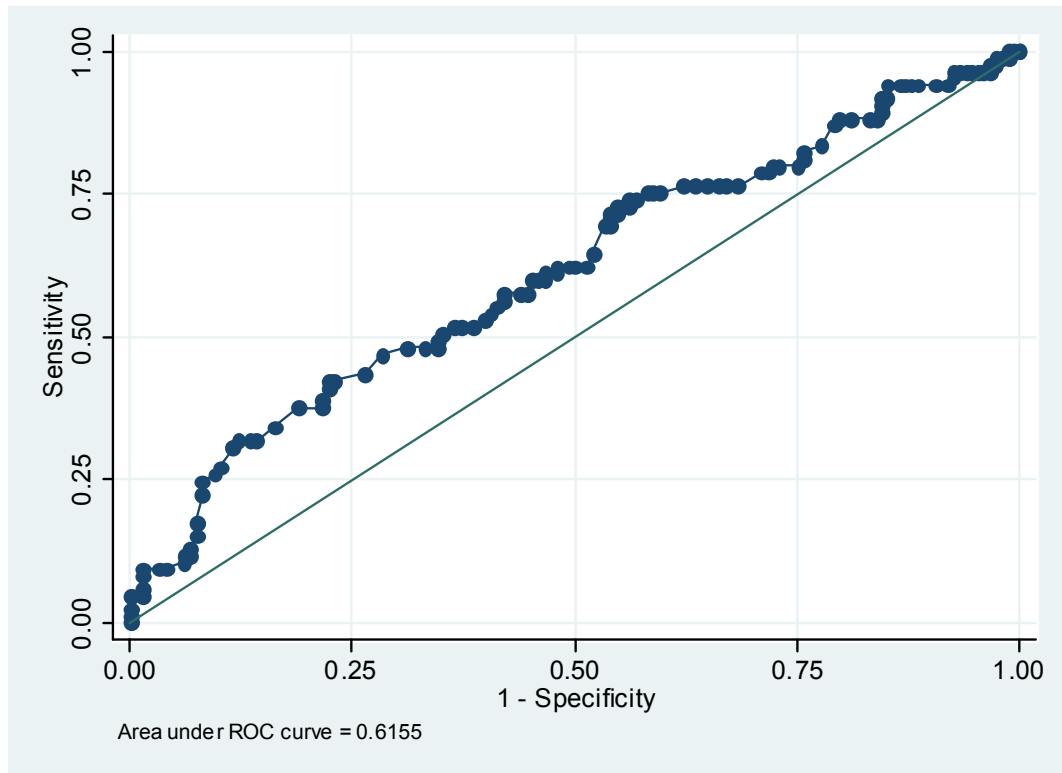


Figure 7-6. ROC Curve in the Memphis Testing Dataset

Next, we used if-then decision tree rules to build a model of AMD. Several combinations of susceptibility factors were not observed in the training dataset (Figure 7-7), and the 3 individuals in the testing dataset who carried these combinations could not be classified. The ratio of cases to controls in each cell had to exceed 1.94 (the ratio of cases to controls in the training dataset) for a combination to be classified as affected. Table 7-9 shows the if-then rules.

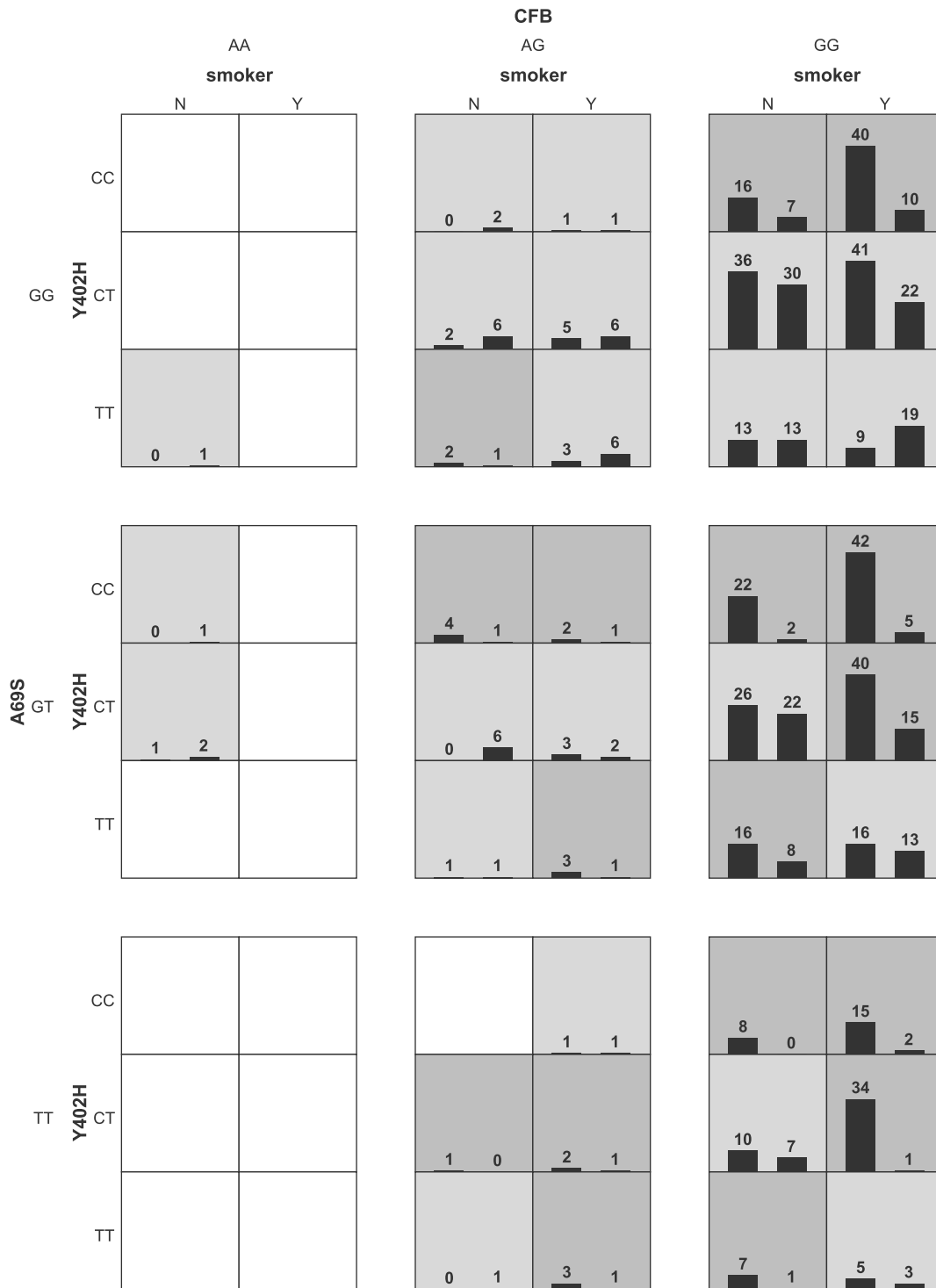


Figure 7-7. Counts of Cases and Controls for Each Susceptibility Combination in the Complete Vanderbilt/Duke Training Dataset. Cases are plotted on the left, controls on the right of each square. Combinations leading to a model call of "affected" are shaded dark gray, and those resulting in a model call of "normal" are shaded light gray. White cells occurred when no one in the training dataset had that particular combination of susceptibility factors.

Table 7-9. If-Then Decision Tree Rules Developed in the Full Vanderbilt/Duke Training Dataset. *If-then rules that were different when the full Vanderbilt/Duke Dataset was used as the training dataset, compared to the rules in Table 7-5, when only half the dataset was used.

```

IF smoker = N AND Y402H = CC AND R32Q = AA AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CC AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CC AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = N AND Y402H = CC AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Affected. *
IF smoker = N AND Y402H = CC AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = N AND Y402H = CT AND R32Q = AA AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = CT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = TT AND R32Q = AA AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Affected. *
IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = TT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = N AND Y402H = TT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CC AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = CC AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CC AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Normal . *
IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CC AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = CT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = CT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = GT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = TT AND R32Q = AG AND A69S = TT THEN CLASSI FY AS Affected.
IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = GG THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = GT THEN CLASSI FY AS Normal .
IF smoker = Y AND Y402H = TT AND R32Q = GG AND A69S = TT THEN CLASSI FY AS Normal . *

```

In this instance, the if-then decision tree rules outperformed the logistic regression model by correctly classifying 62% of the Memphis dataset (Table 7-10). However, this was still a poor model, as cases were correctly identified less than half the time.

Table 7-10. Evaluation of the If-Then Decision Tree Rules in the Memphis Testing Dataset.

| Sensitivity | Specificity | Overall Correctly Classified |
|--------------------|--------------------|-------------------------------------|
| 39/85 (46%) | 106/148 (72%) | 145/233 (62%) |

Discussion

Even though several AMD susceptibility factors that replicate in numerous independent populations have been identified, we were still unable to develop a model of AMD that achieves accuracy rates high enough to have clinical utility. Several factors are likely contributing to the low accuracy of the models.

Sample size

In the analysis using half the Vanderbilt/Duke dataset to create the logistic model and then the other half of the dataset to make predictions, the sample size was probably too small to accurately estimate the true population value of the coefficient estimates, as evidenced by the unimpressive p-values for CFH Y402H, LOC387715 A69S, and smoking. Even when we used the complete Vanderbilt/Duke dataset to build the model and then predicted in the Memphis samples, there were 16 empty cells and only 17 of 54 cells had at least 10 observations in the complete Vanderbilt/Duke training dataset.

Addition or subtraction of even one individual in the training dataset could therefore change the decision rule, leading to an unstable model with decreased accuracy.

Ascertainment Scheme

The Vanderbilt/Duke dataset was ascertained in retina clinics and consists mostly of individuals with severe grade 5 AMD and the cleanest grade 1 controls. In contrast, the Memphis dataset represents the general population. Therefore, we were building the model on the extremes of the “AMD distribution”, and then asking the model to classify everyone along the continuum. This likely explains why the predictive accuracy in the Vanderbilt/Duke testing population was higher than the accuracy in the Memphis cohort, and suggests that if we intend to apply a model of AMD to the general population we should build that model in a population-based cohort.

Missing Variables

While there is some evidence that increased body mass index (BMI) raises the risk for AMD, BMI was not included in either the logistic regression or if-then decision tree models because it shows the weakest association with AMD and we did not want to further stratify the datasets. For similar reasons, we excluded APOE, the CFHL1/CFHL3 deletion, and an interaction term for LOC387715 A69S and cigarette smoking. Variants in complement component 3 were not included, because this association was only very recently discovered. Finally, there may be risk or protective factors for AMD that have yet to be identified and including these in the models could greatly raise their accuracy.

In conclusion, at present there is no predictive test for AMD that can be accurately applied to the general population. Large cohorts representative of the population will be needed to create such a test, assuming that all factors that strongly influence AMD susceptibility can be identified.

CHAPTER VIII

CONCLUSIONS AND FUTURE DIRECTIONS

Despite its underlying complexity, great strides have been made in unraveling the genetic causes of AMD. With different polymorphisms in the same gene acting either to increase or decrease risk, genetic associations varying by ethnicity, and gene-environment interactions, AMD certainly fits the model for a complex disease.

For example, not only do polymorphisms in CFH elevate risk, certain haplotypes in CFH are protective for AMD. Deletion of the CFH-like genes CFHL1 and CFHL3 segregating with one of these haplotypes seems to be at the root of this protective effect, in some, but not all cases. In the present study, we confirmed the presence of a deletion and its inverse association with AMD in homozygotes, but did not determine whether the deletion is protective in the heterozygous state. Also, we have yet to map the breakpoints of the deletion in our sample to ensure that all individuals carry the same deletion and that it matches the endpoints first described by Hughes et al. Variants on protective haplotypes that don't carry the deletion still need to be pinpointed.

The function of the 5 CFH-like genes is unknown, but they are strongly suspected to have similar roles to CFH in immune response given the high sequence identity between the genes. Though it was beyond the scope of this project, a thorough screening of these genes might unearth new AMD susceptibility variants.

Most gene mapping studies focus on polymorphisms that raise disease susceptibility, but risk is only one side of the story. Using a candidate gene approach focused on the

complement pathways, a protective variant R32Q in CFB was identified, and the possibility of a second protective variant in the adjacent CC2 gene was raised, though subsequent reports could not distinguish between a causal effect of CC2 E318D or an association due to linkage disequilibrium with nearby CFB polymorphisms. Hoping to clarify the issue, in the present study we dissected this region of chromosome 6, and using conditional analyses and likelihood ratio testing, we observed strong association of CFB R32Q and modest association of E318D CC2. Given the modest association observed in our study, replication of the CC2 association is still needed to clinch the argument that it is an independent AMD modulator.

To further complicate matters, genetic risk factors must be placed in the context of the environment. Smoking has long been known to raise the odds of AMD, but it was recently discovered that smoking and LOC387715 A69S in LOC387715 act synergistically to produce higher odds of AMD together than would be expected from the odds of each factor alone. However, some controversy surrounds whether the LOC387715 A69S change in LOC387715 is the “causal” AMD variant on chromosome 10, or if a promoter SNP in the nearby HTRA1 gene has that distinction. As the SNPs are in very strong linkage disequilibrium, teasing this apart remains an ongoing struggle. Furthermore, the serine protease function of HTRA1 has been fairly well characterized, but the biological role of LOC387715 has yet to be elucidated. While these questions were beyond the scope of the current proposal, definitive answers will be needed to advance the field.

Lastly, the search for novel AMD loci continues. On chromosome 16, we have identified strong linkage to CACNG3 and association of polymorphisms in both the

family-based and case-control datasets. We have postulated that duplication of the 5' upstream sequence and exon 1 of CACNG3 may be the source of these results, but this has not been confirmed. Efforts to increase SNP density across the 5' end of the gene and to genotype the duplication are ongoing.

Putting all we have learned about AMD together, can we now predict who is likely to be affected? Based on the poor predictive accuracy of the AMD models developed in the present study when applied to prospective cohorts, the answer seems to be “not yet”. Rather than be discouraged, this should prompt us to gather larger sample sizes so that regression coefficient estimates and decision tree rules will be more stable and to continue refining our knowledge of the biological mechanisms underlying the pathophysiology of AMD so that more accurate models can be developed.

In conclusion, while some of the secrets of the genetic etiology of AMD have been revealed, much of the puzzle remains to be solved. Future work will focus on confirming recently proposed risk variants, identifying the remaining AMD susceptibility factors, and determining how these genes and the environment act biologically to promote disease. Perhaps someday this knowledge will be translated into new therapies that improve the quality of life for AMD patients or preventative measures that benefit the general population.

APPENDIX

Table A-1 Description of SNPs in the Initial Chr. 16p12 Screening Analyses. A “.” for SNPs in the case-control dataset indicates that the SNP was only genotyped in the family-based dataset as part of the Illumina Genome Screen. This table contains a column for % of sample genotyped, rather than genotyping efficiency, because additional samples were added to the dataset while the screening was occurring and not all SNPs were genotyped on the most current dataset.

| SNP | Position | Minor Allele | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| | | | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS2024495 | 10317112 | C | 97 | 0.01 | 0.38 | . | . | . | . | . |
| RS10221 | 10484796 | A | 99 | 0.42 | 0.49 | . | . | . | . | . |
| RS2251984 | 10695607 | A | 90 | 0.76 | 0.48 | 91 | 0.65 | 1.00 | 0.48 | 0.50 |
| RS887864 | 11066386 | C | 98 | 0.58 | 0.36 | . | . | . | . | . |
| RS741175 | 11067186 | C | 98 | 0.53 | 0.42 | . | . | . | . | . |
| RS918738 | 11347180 | C | 100 | 0.26 | 0.48 | . | . | . | . | . |
| RS3743582 | 11557826 | A | 91 | 0.50 | 0.50 | 89 | 0.41 | 0.39 | 0.97 | 0.47 |
| RS1035579 | 12438660 | T | 100 | 0.03 | 0.49 | . | . | . | . | . |
| RS1002970 | 13339834 | G | 99 | 0.41 | 0.31 | . | . | . | . | . |
| RS734826 | 13397881 | T | 100 | 0.36 | 0.41 | . | . | . | . | . |
| RS1001937 | 13439013 | G | 100 | 0.06 | 0.41 | . | . | . | . | . |
| RS928963 | 13523732 | A | 90 | 0.16 | 0.37 | 89 | 0.58 | 0.44 | 0.92 | 0.41 |
| RS1158123 | 13616366 | G | 99 | 0.76 | 0.47 | . | . | . | . | . |
| RS11649492 | 13916919 | C | 92 | 0.33 | 0.37 | 91 | 0.59 | 0.54 | 1.00 | 0.36 |
| RS7200272 | 14572980 | T | 93 | 0.45 | 0.31 | 93 | 0.18 | 0.06 | 0.74 | 0.29 |
| RS2384933 | 15723647 | T | 100 | 0.00 | 0.34 | . | . | . | . | . |
| RS212090 | 16143505 | T | 99 | 0.79 | 0.42 | . | . | . | . | . |
| RS1597987 | 16791812 | G | 90 | 0.38 | 0.48 | 94 | 0.79 | 0.80 | 0.31 | 0.50 |
| RS1472426 | 17099935 | T | 98 | 0.01 | 0.45 | . | . | . | . | . |
| RS936347 | 17178296 | A | 100 | 0.67 | 0.49 | . | . | . | . | . |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS1457907 | 17253737 | T | 99 | 0.61 | 0.43 | . | . | . | . | . |
| RS1379652 | 17304099 | G | 100 | 0.35 | 0.49 | . | . | . | . | . |
| RS7195703 | 17383146 | C | 92 | 0.74 | 0.43 | 92 | 0.45 | 0.47 | 0.87 | 0.46 |
| RS1389504 | 17841806 | A | 99 | 0.80 | 0.40 | . | . | . | . | . |
| RS757189 | 17895804 | A | 98 | 0.02 | 0.46 | . | . | . | . | . |
| RS904821 | 18747173 | A | 100 | 0.01 | 0.45 | . | . | . | . | . |
| RS1544357 | 19117111 | A | 100 | 0.60 | 0.33 | . | . | . | . | . |
| RS724307 | 19173482 | G | 97 | 0.36 | 0.42 | . | . | . | . | . |
| RS2023762 | 19184098 | G | 99 | 0.97 | 0.43 | . | . | . | . | . |
| RS179209 | 19214631 | A | 100 | 0.17 | 0.50 | . | . | . | . | . |
| RS179219 | 19220873 | T | 99 | 0.09 | 0.47 | . | . | . | . | . |
| RS875648 | 19570770 | C | 99 | 0.09 | 0.41 | . | . | . | . | . |
| RS227761 | 19768806 | G | 99 | 0.15 | 0.49 | . | . | . | . | . |
| RS727590 | 19953169 | G | 97 | 0.64 | 0.24 | . | . | . | . | . |
| RS151328 | 20556203 | A | 88 | 0.31 | 0.23 | 70 | 0.41 | 0.72 | 0.43 | 0.19 |
| RS3785080 | 20659737 | A | 88 | 0.78 | 0.26 | 71 | 0.75 | 1.00 | 0.48 | 0.30 |
| RS2107232 | 20731822 | A | 87 | 0.73 | 0.26 | 70 | 0.84 | 0.96 | 0.49 | 0.30 |
| RS1978091 | 20794271 | G | 87 | 0.81 | 0.26 | 70 | 0.79 | 1.00 | 0.49 | 0.30 |
| RS8059938 | 20853437 | C | 87 | 0.34 | 0.26 | 54 | 0.79 | 1.00 | 0.56 | 0.29 |
| RS330150 | 20958710 | C | 86 | 0.50 | 0.10 | 55 | 0.30 | 0.22 | 1.00 | 0.08 |
| RS861424 | 20986194 | G | 88 | 0.94 | 0.42 | 70 | 1.00 | 0.45 | 0.29 | 0.40 |
| RS2031077 | 21145237 | T | 100 | 0.01 | 0.32 | . | . | . | . | . |
| RS741720 | 21147979 | T | 99 | 1.00 | 0.43 | . | . | . | . | . |
| RS2733910 | 21192782 | T | 88 | 1.00 | 0.31 | 70 | 0.23 | 0.48 | 0.35 | 0.27 |
| RS1055740 | 21560221 | A | 88 | 0.86 | 0.15 | 71 | 0.10 | 0.72 | 0.02 | 0.15 |
| RS215901 | 21637950 | C | 88 | 0.45 | 0.30 | 70 | 0.18 | 0.30 | 0.49 | 0.31 |
| RS2968403 | 21855288 | A | 93 | 1.00 | 0.14 | 87 | 1.00 | 1.00 | 1.00 | 0.09 |
| HCV1395277 | 21993631 | A | 88 | 0.29 | 0.13 | 71 | 1.00 | 0.69 | 0.70 | 0.12 |
| RS6497580 | 22139927 | T | 87 | 0.82 | 0.35 | 70 | 0.27 | 0.15 | 0.98 | 0.32 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS2926362 | 22237539 | T | 88 | 0.16 | 0.35 | 70 | 0.99 | 0.64 | 0.36 | 0.35 |
| RS2239331 | 22731955 | A | 86 | 0.92 | 0.49 | 55 | 0.18 | 0.21 | 0.72 | 0.39 |
| RS169660 | 22748205 | A | 86 | 0.38 | 0.34 | 54 | 0.88 | 0.70 | 0.89 | 0.38 |
| RS7198577 | 22766504 | A | 87 | 1.00 | 0.32 | 55 | 0.59 | 0.99 | 0.40 | 0.36 |
| RS208965 | 22769836 | C | 100 | 1.00 | 0.50 | . | . | . | . | . |
| RS208626 | 22835388 | G | 86 | 0.41 | 0.44 | 53 | 0.96 | 1.00 | 0.90 | 0.46 |
| RS1011463 | 22957621 | T | 88 | 0.76 | 0.39 | 70 | 0.02 | 0.01 | 0.89 | 0.32 |
| RS1858799 | 23022885 | C | 100 | 0.90 | 0.37 | . | . | . | . | . |
| RS715368 | 23048647 | C | 88 | 1.00 | 0.38 | 71 | 0.04 | 0.01 | 0.84 | 0.32 |
| RS5723 | 23134288 | G | 87 | 0.18 | 0.21 | 71 | 0.85 | 0.67 | 0.98 | 0.20 |
| RS886113 | 23238208 | A | 88 | 0.71 | 0.36 | 70 | 0.81 | 0.75 | 1.00 | 0.36 |
| RS238547 | 23267700 | T | 92 | 0.02 | 0.38 | 90 | 0.95 | 0.22 | 0.03 | 0.44 |
| RS2303153 | 23297702 | C | 86 | 0.89 | 0.49 | 71 | 0.47 | 1.00 | 0.14 | 0.49 |
| RS4968008 | 23397854 | T | 87 | 0.59 | 0.21 | 70 | 0.91 | 0.62 | 0.16 | 0.19 |
| RS3809682 | 23632254 | G | 86 | 0.78 | 0.33 | 55 | 0.53 | 0.33 | 0.83 | 0.35 |
| RS194790 | 23699658 | G | 100 | 0.85 | 0.49 | . | . | . | . | . |
| RS2023671 | 23757340 | G | 88 | 0.46 | 0.25 | 70 | 0.57 | 0.88 | 0.56 | 0.25 |
| RS2188356 | 23839214 | C | 83 | 0.58 | 0.44 | 69 | 1.00 | 0.32 | 0.09 | 0.40 |
| RS916677 | 23910339 | T | 88 | 0.79 | 0.42 | 70 | 0.11 | 0.10 | 0.73 | 0.43 |
| RS1490754 | 23922567 | C | 94 | 0.82 | 0.38 | 93 | 0.98 | 0.93 | 0.77 | 0.41 |
| RS405322 | 24022987 | T | 86 | 0.61 | 0.49 | 70 | 0.55 | 0.82 | 0.44 | 0.47 |
| RS12448206 | 24096007 | A | 86 | 0.59 | 0.22 | 71 | 1.00 | 1.00 | 1.00 | 0.23 |
| RS411103 | 24137704 | A | 99 | 0.92 | 0.45 | . | . | . | . | . |
| RS2238500 | 24192709 | G | 97 | 0.01 | 0.43 | 97 | 0.72 | 0.70 | 1.00 | 0.44 |
| RS991911 | 24198554 | A | 100 | 0.03 | 0.38 | . | . | . | . | . |
| RS757200 | 24227195 | T | 93 | 0.00 | 0.29 | 91 | 0.04 | 0.03 | 0.63 | 0.30 |
| RS2238518 | 24247639 | C | 92 | 1.00 | 0.44 | 91 | 0.19 | 0.03 | 0.42 | 0.44 |
| RS2189290 | 24264034 | C | 87 | 0.39 | 0.39 | 71 | 0.54 | 0.60 | 0.04 | 0.32 |
| RS12596694 | 24455938 | C | 86 | 0.66 | 0.13 | 70 | 0.40 | 1.00 | 0.25 | 0.13 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS2033214 | 24485021 | G | 86 | 0.30 | 0.13 | 70 | 0.97 | 1.00 | 0.94 | 0.12 |
| RS11641483 | 24500126 | G | 88 | 0.12 | 0.18 | 92 | 0.73 | 0.44 | 0.70 | 0.16 |
| RS1549239 | 24732986 | T | 86 | 0.49 | 0.30 | 54 | 0.68 | 0.70 | 1.00 | 0.27 |
| RS1011078 | 24801178 | G | 99 | 1.00 | 0.48 | . | . | . | . | . |
| RS12927084 | 24806905 | A | 86 | 0.81 | 0.46 | 70 | 0.41 | 0.47 | 0.83 | 0.49 |
| RS6497763 | 24831156 | C | 92 | 0.17 | 0.49 | 89 | 0.96 | 0.37 | 0.11 | 0.48 |
| RS1035946 | 24881927 | G | 86 | 0.67 | 0.20 | 70 | 0.97 | 0.92 | 0.61 | 0.19 |
| RS4787690 | 25005001 | G | 88 | 0.80 | 0.47 | 70 | 0.74 | 0.49 | 0.73 | 0.49 |
| RS874562 | 25010021 | A | 99 | 0.81 | 0.44 | . | . | . | . | . |
| RS12925518 | 25098344 | A | 86 | 0.66 | 0.25 | 70 | 1.00 | 0.24 | 0.04 | 0.23 |
| RS2345122 | 25156195 | A | 90 | 0.44 | 0.48 | 86 | 1.00 | 0.29 | 0.08 | 0.50 |
| RS8049535 | 25179171 | G | 87 | 0.53 | 0.37 | 69 | 0.84 | 0.69 | 0.99 | 0.33 |
| RS2157857 | 25420748 | C | 100 | 0.87 | 0.45 | 70 | 0.11 | 0.56 | 0.08 | 0.45 |
| RS1022455 | 25436891 | T | 99 | 0.94 | 0.47 | . | . | . | . | . |
| RS205162 | 25504138 | T | 99 | 0.32 | 0.18 | . | . | . | . | . |
| RS2966220 | 25595211 | A | 86 | 1.00 | 0.40 | 54 | 0.20 | 0.45 | 0.27 | 0.44 |
| RS4520838 | 25701169 | T | 86 | 0.51 | 0.35 | 70 | 0.22 | 0.25 | 0.73 | 0.38 |
| HCV365765 | 25795966 | C | 84 | 0.43 | 0.35 | 69 | 0.37 | 0.52 | 0.66 | 0.34 |
| RS7190163 | 25907847 | C | 87 | 0.83 | 0.29 | 71 | 0.96 | 0.72 | 0.38 | 0.28 |
| RS7188016 | 26007978 | T | 87 | 0.02 | 0.47 | 70 | 0.06 | 0.06 | 0.85 | 0.47 |
| HCV504442 | 26057566 | T | 88 | 0.02 | 0.29 | 71 | 0.42 | 0.32 | 1.00 | 0.26 |
| RS730015 | 26260645 | G | 86 | 0.07 | 0.44 | 70 | 0.79 | 0.63 | 0.90 | 0.41 |
| RS2078274 | 26508367 | A | 100 | 0.72 | 0.32 | . | . | . | . | . |
| RS739480 | 26524186 | C | 100 | 0.32 | 0.20 | . | . | . | . | . |
| RS237131 | 26547907 | A | 99 | 0.85 | 0.36 | . | . | . | . | . |
| RS2042347 | 26636056 | G | 86 | 0.90 | 0.45 | 55 | 0.17 | 0.01 | 0.14 | 0.41 |
| RS723876 | 26793588 | C | 86 | 0.26 | 0.35 | 69 | 0.36 | 0.92 | 0.16 | 0.37 |
| RS11823 | 26987807 | G | 88 | 0.61 | 0.28 | 71 | 0.62 | 0.94 | 0.23 | 0.26 |
| RS4787929 | 27063535 | G | 86 | 0.58 | 0.26 | 69 | 0.38 | 0.39 | 0.93 | 0.24 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS4499236 | 27161632 | T | 87 | 0.00 | 0.44 | 70 | 0.69 | 0.82 | 0.80 | 0.48 |
| RS2057768 | 27229596 | T | 100 | 1.00 | 0.28 | . | . | . | . | . |
| RS3024548 | 27262032 | G | 91 | 0.12 | 0.45 | 86 | 0.30 | 0.97 | 0.04 | 0.47 |
| RS8832 | 27283288 | A | 100 | 0.22 | 0.44 | . | . | . | . | . |
| RS3093291 | 27346984 | C | 98 | 0.60 | 0.33 | . | . | . | . | . |
| RS232073 | 27421360 | G | 88 | 0.19 | 0.16 | 70 | 0.13 | 0.31 | 0.39 | 0.14 |
| RS4787976 | 27513198 | C | 88 | 0.44 | 0.19 | 71 | 0.32 | 0.96 | 0.12 | 0.17 |
| RS1017575 | 27597669 | G | 88 | 0.40 | 0.16 | 71 | 0.05 | 0.19 | 0.27 | 0.14 |
| RS8054700 | 27694169 | T | 92 | 0.29 | 0.15 | 89 | 0.68 | 0.76 | 0.10 | 0.15 |
| RS772859 | 27701027 | G | 83 | 0.65 | 0.38 | 53 | 0.88 | 0.62 | 0.78 | 0.38 |
| RS755297 | 27713895 | C | 90 | 0.03 | 0.16 | 92 | 0.81 | 0.68 | 0.22 | 0.18 |
| RS1559167 | 27726752 | T | 94 | 0.14 | 0.43 | 91 | 0.77 | 0.76 | 1.00 | 0.44 |
| RS1644609 | 27734413 | A | 86 | 1.00 | 0.41 | 88 | 0.95 | 0.87 | 1.00 | 0.40 |
| RS713547 | 27739590 | C | 87 | 0.08 | 0.13 | 70 | 0.74 | 0.20 | 0.02 | 0.14 |
| RS1644618 | 27757739 | T | 92 | 0.89 | 0.29 | 93 | 0.68 | 0.60 | 1.00 | 0.27 |
| RS6498040 | 27776664 | T | 89 | 0.31 | 0.41 | 72 | 0.09 | 0.33 | 0.13 | 0.40 |
| RS4788003 | 27806901 | T | 93 | 1.00 | 0.34 | 92 | 0.32 | 0.68 | 0.30 | 0.36 |
| RS2385008 | 27841534 | A | 87 | 0.68 | 0.37 | 70 | 0.73 | 0.78 | 0.95 | 0.34 |
| RS7194904 | 27874886 | A | 92 | 0.81 | 0.45 | 93 | 0.47 | 0.32 | 0.94 | 0.44 |
| RS4788017 | 27884330 | G | 88 | 0.39 | 0.23 | 70 | 0.76 | 0.96 | 0.43 | 0.24 |
| RS9941112 | 27885980 | A | 86 | 0.64 | 0.35 | 70 | 0.10 | 0.11 | 0.65 | 0.37 |
| RS1008409 | 27901325 | C | 93 | 0.14 | 0.32 | 93 | 0.95 | 0.72 | 0.80 | 0.33 |
| RS11074888 | 27930557 | C | 92 | 0.51 | 0.39 | 93 | 0.36 | 0.19 | 0.86 | 0.41 |
| RS1476507 | 27932571 | C | 88 | 0.15 | 0.39 | 70 | 0.75 | 0.91 | 0.78 | 0.41 |
| RS205418 | 27980171 | G | 89 | 0.83 | 0.26 | 71 | 0.02 | 0.01 | 0.84 | 0.26 |
| RS12919673 | 28072410 | T | 87 | 0.34 | 0.31 | 71 | 0.22 | 0.09 | 0.94 | 0.28 |
| RS3922800 | 28107380 | C | 88 | 0.26 | 0.31 | 68 | 0.14 | 0.08 | 1.00 | 0.28 |
| RS4453501 | 28167502 | A | 89 | 0.35 | 0.31 | 71 | 0.19 | 0.08 | 1.00 | 0.27 |
| RS2726040 | 28238291 | G | 93 | 1.00 | 0.40 | 92 | 0.88 | 0.99 | 0.88 | 0.40 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|--------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS151233 | 28413929 | T | 91 | 1.00 | 0.14 | 82 | 0.02 | 0.03 | 0.37 | 0.13 |
| APOB48R-9270 | 28415145 | G | 88 | 0.21 | 0.45 | 85 | 1.00 | 0.58 | 0.36 | 0.39 |
| RS40834 | 28417894 | G | 90 | 0.10 | 0.49 | 80 | 0.75 | 0.68 | 0.16 | 0.45 |
| RS40835 | 28417956 | G | 93 | 0.18 | 0.09 | 86 | 0.60 | 0.17 | 0.65 | 0.07 |
| RS181207 | 28421031 | A | 95 | 1.00 | 0.37 | 95 | 0.46 | 0.29 | 0.84 | 0.32 |
| RS151228 | 28470527 | T | 88 | 0.07 | 0.41 | 71 | 0.40 | 0.06 | 0.28 | 0.46 |
| RS12443881 | 28749278 | T | 89 | 0.01 | 0.42 | 87 | 0.88 | 0.40 | 0.25 | 0.38 |
| HCV105407 | 28753086 | T | 87 | 0.00 | 0.22 | 70 | 0.44 | 0.80 | 0.37 | 0.25 |
| RS7193733 | 28782983 | G | 91 | 0.02 | 0.42 | 85 | 0.36 | 0.19 | 0.72 | 0.39 |
| RS2904880 | 28851897 | C | 95 | 0.02 | 0.37 | 88 | 1.00 | 0.74 | 0.64 | 0.32 |
| RS2070962 | 28857933 | T | 92 | 0.00 | 0.36 | 91 | 0.99 | 0.67 | 0.41 | 0.37 |
| RS11150675 | 28888032 | A | 93 | 0.00 | 0.36 | 91 | 0.77 | 0.31 | 0.41 | 0.38 |
| RS4077347 | 28944416 | G | 91 | 0.30 | 0.50 | 85 | 0.17 | 0.31 | 0.39 | 0.50 |
| RS1646129 | 29048189 | A | 99 | 0.34 | 0.34 | . | . | . | . | . |
| RS1642026 | 29056066 | C | 92 | 0.02 | 0.37 | 85 | 0.16 | 0.93 | 0.02 | 0.36 |
| RS252342 | 29094079 | A | 92 | 0.93 | 0.49 | 86 | 0.22 | 0.34 | 0.63 | 0.50 |
| RS252313 | 29121627 | A | 93 | 0.77 | 0.39 | 87 | 0.99 | 0.21 | 0.08 | 0.34 |
| RS10871481 | 29191613 | T | 90 | 0.22 | 0.37 | 85 | 0.57 | 0.91 | 0.48 | 0.37 |
| RS1531974 | 29224597 | A | 88 | 0.64 | 0.33 | 70 | 0.45 | 0.99 | 0.25 | 0.31 |
| RS871887 | 29226120 | A | 92 | 0.71 | 0.30 | 84 | 0.00 | 0.02 | 0.05 | 0.27 |
| RS2171223 | 29234054 | G | 90 | 0.14 | 0.34 | 85 | 0.03 | 0.06 | 0.26 | 0.32 |
| RS8054172 | 29563365 | C | 91 | 0.26 | 0.34 | 92 | 0.59 | 0.91 | 0.49 | 0.40 |
| RS1364184 | 29563711 | T | 92 | 1.00 | 0.07 | 86 | 0.49 | 0.42 | 1.00 | 0.07 |
| RS11859842 | 29568718 | A | 92 | 0.12 | 0.43 | 85 | 0.09 | 0.03 | 0.97 | 0.47 |
| RS11150564 | 29574237 | A | 91 | 0.16 | 0.40 | 86 | 0.53 | 0.81 | 0.53 | 0.46 |
| RS4788172 | 29575754 | A | 93 | 1.00 | 0.07 | 92 | 0.92 | 1.00 | 1.00 | 0.06 |
| RS3764276 | 29580704 | C | 91 | 0.10 | 0.09 | 85 | 0.65 | 0.47 | 1.00 | 0.08 |
| RS2071420 | 29582324 | C | 96 | 0.29 | 0.10 | 96 | 1.00 | 1.00 | 1.00 | 0.09 |
| RS1050881 | 29583429 | T | 88 | 1.00 | 0.07 | 57 | 0.94 | 0.26 | 0.03 | 0.07 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|------------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS6565169 | 29600193 | C | 90 | 0.01 | 0.44 | 93 | 0.22 | 0.78 | 0.09 | 0.37 |
| RS648559 | 29648146 | C | 93 | 0.14 | 0.19 | 86 | 0.48 | 0.37 | 1.00 | 0.20 |
| RS235659 | 29696420 | C | 92 | 0.66 | 0.31 | 86 | 0.65 | 0.39 | 0.80 | 0.33 |
| HCV2851897 | 29727984 | A | 88 | 0.97 | 0.14 | 71 | 1.00 | 0.95 | 1.00 | 0.16 |
| RS1057451 | 29740989 | A | 92 | 0.97 | 0.15 | 86 | 1.00 | 0.93 | 1.00 | 0.17 |
| RS4788186 | 29748726 | T | 92 | 0.41 | 0.34 | 85 | 0.15 | 0.21 | 0.54 | 0.38 |
| RS3815822 | 29779862 | C | 86 | 0.58 | 0.40 | 54 | 1.00 | 1.00 | 1.00 | 0.44 |
| RS4548895 | 29831011 | A | 91 | 0.38 | 0.37 | 85 | 0.91 | 0.79 | 1.00 | 0.38 |
| RS12716972 | 29844155 | A | 91 | 0.93 | 0.43 | 86 | 0.64 | 0.93 | 0.54 | 0.44 |
| RS6565173 | 29881668 | T | 95 | 0.40 | 0.45 | 87 | 0.97 | 0.74 | 0.75 | 0.46 |
| RS11901 | 29891571 | C | 100 | 0.47 | 0.45 | 92 | 0.56 | 0.71 | 0.75 | 0.47 |
| RS11150581 | 29938200 | G | 87 | 0.57 | 0.41 | 70 | 0.81 | 0.36 | 0.03 | 0.41 |
| RS11642740 | 29968156 | C | 88 | 0.41 | 0.42 | 71 | 0.98 | 0.31 | 0.06 | 0.42 |
| RS11860935 | 29988281 | T | 92 | 0.81 | 0.08 | 91 | 1.00 | 0.33 | 0.19 | 0.08 |
| RS8060511 | 30009097 | G | 88 | 0.98 | 0.49 | 86 | 0.30 | 0.39 | 0.69 | 0.50 |
| RS3809624 | 30010303 | G | 89 | 0.37 | 0.30 | 71 | 1.00 | 1.00 | 1.00 | 0.29 |
| RS9924308 | 30062241 | G | 91 | 0.39 | 0.50 | 85 | 0.45 | 0.55 | 0.71 | 0.49 |
| RS7202714 | 30085308 | A | 90 | 0.91 | 0.32 | 85 | 0.04 | 0.10 | 0.23 | 0.31 |
| RS1132812 | 30105652 | T | 86 | 0.03 | 0.47 | 90 | 0.84 | 0.96 | 0.55 | 0.47 |
| RS11862806 | 30271572 | C | 88 | 1.00 | 0.35 | 71 | 0.27 | 0.44 | 0.49 | 0.37 |
| RS12921440 | 30316266 | G | 86 | 0.76 | 0.38 | 88 | 0.10 | 0.46 | 0.09 | 0.37 |
| RS3574 | 30318883 | G | 100 | 0.69 | 0.38 | 69 | 0.09 | 0.33 | 0.14 | 0.39 |
| RS4787645 | 30364851 | A | 87 | 0.20 | 0.41 | 70 | 0.45 | 1.00 | 0.12 | 0.37 |
| RS2230433 | 30425542 | G | 95 | 0.94 | 0.28 | 88 | 0.08 | 0.05 | 1.00 | 0.29 |
| RS12918327 | 30534117 | A | 88 | 0.11 | 0.23 | 70 | 0.61 | 0.72 | 0.91 | 0.24 |
| RS885107 | 30580220 | G | 87 | 0.10 | 0.26 | 69 | 0.51 | 0.85 | 0.49 | 0.28 |
| HCV1115064 | 30699890 | G | 88 | 0.83 | 0.15 | 71 | 0.52 | 0.98 | 0.35 | 0.18 |
| RS1046276 | 30822127 | A | 88 | 0.93 | 0.35 | 71 | 0.08 | 0.72 | 0.02 | 0.34 |
| RS3751855 | 30998710 | G | 87 | 0.43 | 0.39 | 70 | 0.55 | 0.63 | 0.81 | 0.37 |

| | | | Family 345vs12 | | | Case-Control 345vs12 | | | | |
|-----------|----------|--------------|-----------------------|---------|------|-----------------------|---------|-----------|--------------|------|
| SNP | Position | Minor Allele | % of Sample Genotyped | HWE All | MAF | % of Sample Genotyped | HWE All | HWE Cases | HWE Controls | MAF |
| RS14235 | 31029294 | T | 86 | 0.52 | 0.37 | 54 | 0.53 | 0.74 | 0.67 | 0.35 |
| RS889551 | 31167923 | T | 98 | 0.36 | 0.34 | . | . | . | . | . |
| RS13143 | 31396534 | A | 99 | 0.20 | 0.24 | . | . | . | . | . |
| RS1534507 | 31567928 | G | 99 | 0.10 | 0.23 | . | . | . | . | . |
| RS9929259 | 31718287 | C | 86 | 0.19 | 0.28 | 53 | 0.15 | 0.20 | 0.70 | 0.30 |

REFERENCES

1. Abecasis GR et al (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat.Genet.* 30 (1):97-101.
2. Abecasis GR et al (2004) Age-related macular degeneration: a high-resolution genome scan for susceptibility loci in a population enriched for late-stage disease. *Am.J.Hum.Genet.* 74 (3):482-494.
3. Alexander JJ and Quigg RJ (2007) The simple design of complement factor H: Looks can be deceiving. *Mol.Immunol.* 44 (1-3):123-132.
4. Anderson DH et al (2002) A role for local inflammation in the formation of drusen in the aging eye. *Am.J.Ophthalmol.* 134 (3):411-431.
5. Anderson DH et al (2001) Local cellular sources of apolipoprotein E in the human retina and retinal pigmented epithelium: implications for the process of drusen formation. *Am.J.Ophthalmol.* 131 (6):767-781.
6. AREDS (1999) The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1. *Control Clin.Trials* 20 (6):573-600.
7. Baird PN et al (2004) The epsilon2 and epsilon4 alleles of the apolipoprotein gene are associated with age-related macular degeneration. *Invest Ophthalmol.Vis.Sci.* 45 (5):1311-1315.
8. Barrett JC et al (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21 (2):263-265.
9. Baudouin C et al (1992) Immunohistological study of subretinal membranes in age-related macular degeneration. *Jpn.J.Ophthalmol.* 36 (4):443-451.
10. Berger, JW, Fine, SL, and Maguire, MG. Age-related Macular Degeneration. 1999. St. Louis, Mosby.
Ref Type: Generic
11. Boyles AL et al (2005) Linkage disequilibrium inflates type I error rates in multipoint linkage analysis when parental genotypes are missing. *Hum Hered.* 59 (4):220-227.
12. Cameron DJ et al (2007) HTRA1 variant confers similar risks to geographic atrophy and neovascular age-related macular degeneration. *Cell Cycle* 6 (9):1122-1125.
13. Caprioli J et al (2003) Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T

- polymorphisms are strongly associated with the disease. *Hum.Mol.Genet.* 12 (24):3385-3395.
14. Chen LJ et al (2006) Association of complement factor H polymorphisms with exudative age-related macular degeneration. *Mol.Vis.* 12:1536-1542.
 15. Chung RH, Hauser ER, and Martin ER (2006) The APL test: extension to general nuclear families and haplotypes and examination of its robustness. *Hum.Hered.* 61 (4):189-199.
 16. Clemons TE et al (2005) Risk factors for the incidence of Advanced Age-Related Macular Degeneration in the Age-Related Eye Disease Study (AREDS) AREDS report no. 19. *Ophthalmology* 112 (4):533-539.
 17. Conley YP et al (2005) Candidate gene analysis suggests a role for fatty acid biosynthesis and regulation of the complement system in the etiology of age-related maculopathy. *Hum.Mol.Genet.* 14 (14):1991-2002.
 18. Cottingham RWJr, Idury RM, and Schaffer AA (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* 53:252-263.
 19. Dewan A et al (2006) HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 314 (5801):989-992.
 20. Duvall-Young J, MacDonald MK, and McKechnie NM (1989) Fundus changes in (type II) mesangiocapillary glomerulonephritis simulating drusen: a histopathological report. *Br.J.Ophthalmol.* 73 (4):297-302.
 21. Edwards AO et al (2005) Complement factor H polymorphism and age-related macular degeneration. *Science* 308 (5720):421-424.
 22. Esparza-Gordilla J and Soria JMBAALBJFJdCSR (2004) Genetic and environmental factors influencing the human factor H plasma levels. *Immunogenetics* 56 (2):77-82.
 23. Fisher SA et al (2005) Meta-analysis of genome scans of age-related macular degeneration. *Hum Mol.Genet* 14 (15):2257-2264.
 24. Fisher SA et al (2006) Assessment of the contribution of CFH and chromosome 10q26 AMD susceptibility loci in a Russian population isolate. *Br.J.Ophthalmol.*
 25. Friedman DS et al (2004) Prevalence of age-related macular degeneration in the United States. *Arch.Ophthalmol.* 122 (4):564-572.
 26. Friedman E (1997) A hemodynamic model of the pathogenesis of age-related macular degeneration. *Am.J.Ophthalmol.* 124 (5):677-682.

27. Fuse N et al (2006) Polymorphisms in Complement Factor H and Hemicentin-1 genes in a Japanese population with dry-type age-related macular degeneration. *Am.J.Ophthalmol.* 142 (6):1074-1076.
28. Gabriel SB et al (2002) The structure of haplotype blocks in the human genome. *Science* 296 (5576):2225-2229.
29. Gallaher KT et al (2007) Estimation of macular pigment optical density in the elderly: test-retest variability and effect of optical blur in pseudophakic subjects. *Vision Res.* 47 (9):1253-1259.
30. Gold, B., Merriam, J. C., Zernant, J., Hancox, L. S., Taiber, A. J., Gehrs, K., Cramer, K, Neel, J, Bergeron, J, Barile, G. R., Smith, R. T., the AMA Genetics Clinical Study Group, Hageman, G. S., Dean, M., and Allikmets, R. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. 3-5-2006. nature genetics.
Ref Type: Generic
31. Gotoh N et al (2004) Apolipoprotein E polymorphisms in Japanese patients with polypoidal choroidal vasculopathy and exudative age-related macular degeneration. *Am.J.Ophthalmol.* 138 (4):567-573.
32. Gotoh N et al (2006) No association between complement factor H gene polymorphism and exudative age-related macular degeneration in Japanese. *Hum.Genet.* 120 (1):139-143.
33. Gottfredsdottir MS et al (1999) Age related macular degeneration in monozygotic twins and their spouses in Iceland. *Acta Ophthalmol.Scand.* 77 (4):422-425.
34. Grizzard SW, Arnett D, and Haag SL (2003) Twin study of age-related macular degeneration. *Ophthalmic Epidemiol.* 10 (5):315-322.
35. Gudbjartsson DF et al (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat.Genet.* 25 (1):12-13.
36. Gudbjartsson DF et al (2005) Allegro version 2. *Nat.Genet* 37 (10):1015-1016.
37. Hageman GS 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.U.S.A* 102 (20):7227-7232.
38. Hageman GS 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.* 38 (8):592-604.
39. Hageman GS 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.* 20 (6):705-732.
40. Haines JL et al (2005) Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308 (5720):419-421.
 41. Hammond CJ et al (2002) Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology* 109 (4):730-736.
 42. Hauser ER et al (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genet.Epidemiol.* 27 (1):53-63.
 43. Hauser MA et al (2003) Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum.Mol.Genet.* 12 (6):671-677.
 44. Heiba IM et al (1994) Sibling correlations and segregation analysis of age-related maculopathy: the Beaver Dam Eye Study. *Genet Epidemiol.* 11 (1):51-67.
 45. Hughes AE 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.* 38 (10):1173-1177.
 46. Iannaccone A et al (2007) Macular pigment optical density in the elderly: findings in a large biracial Midsouth population sample. *Invest Ophthalmol.Vis.Sci.* 48 (4):1458-1465.
 47. Iyengar SK et al (2004) Dissection of genomewide-scan data in extended families reveals a major locus and oligogenic susceptibility for age-related macular degeneration. *Am.J.Hum.Genet.* 74 (1):20-39.
 48. Jakobsdottir J et al (2005) Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am.J.Hum.Genet.* 77 (3):389-407.
 49. Janeway CA, Jr., Travers P, Walport M, Shlomchik MJ (2001) Immunobiology: the immune system in health and disease. 5 ed. Garland Publishing: New York.
 50. Johnson PT et al (2006) Individuals homozygous for the age-related macular degeneration risk-conferring variant of complement factor H have elevated levels of CRP in the choroid. *Proc.Natl.Acad.Sci.U.S.A* 103 (46):17456-17461.
 51. Jokiranta TS et al (2005) Binding of complement factor H to endothelial cells is mediated by the carboxy-terminal glycosaminoglycan binding site. *Am.J.Pathol.* 167 (4):1173-1181.
 52. Klaver CC et al (1998a) Genetic association of apolipoprotein E with age-related macular degeneration. *Am J Hum Genet* 63 (1):200-206.

53. Klaver CC et al (1998b) Genetic risk of age-related maculopathy. Population-based familial aggregation study. *Arch.Ophthalmol.* 116 (12):1646-1651.
54. Klein BE et al (2001a) Risk of incident age-related eye diseases in people with an affected sibling : The Beaver Dam Eye Study. *Am.J.Epidemiol.* 154 (3):207-211.
55. Klein ML, Mauldin WM, and Stoumbos VD (1994) Heredity and age-related macular degeneration. Observations in monozygotic twins. *Arch.Ophthalmol.* 112 (7):932-937.
56. Klein ML et al (1998) Age-related macular degeneration. Clinical features in a large family and linkage to chromosome 1q. *Arch Ophthalmol* 116 (8):1082-1088.
57. Klein R et al (2007) Cardiovascular disease, its risk factors and treatment, and age-related macular degeneration: Women's Health Initiative Sight Exam ancillary study. *Am.J.Ophthalmol.* 143 (3):473-483.
58. Klein R et al (2001b) Medication use and the 5-year incidence of early age-related maculopathy: the Beaver Dam Eye Study. *Arch.Ophthalmol.* 119 (9):1354-1359.
59. Klein R et al (1997) The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology* 104 (1):7-21.
60. Klein R et al (2006) Prevalence of age-related macular degeneration in 4 racial/ethnic groups in the multi-ethnic study of atherosclerosis. *Ophthalmology* 113 (3):373-380.
61. Klein R, Klein BE, and Linton KL (1992) Prevalence of age-related maculopathy. The Beaver Dam Eye Study. *Ophthalmology* 99 (6):933-943.
62. Klein R et al (2004) The epidemiology of age-related macular degeneration. *Am J Ophthalmol.* 137 (3):486-495.
63. Klein RJ et al (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* 308 (5720):385-389.
64. Laine M et al (2007) Y402H polymorphism of complement factor H affects binding affinity to C-reactive protein. *J.Immunol.* 178 (6):3831-3836.
65. Lau LI et al (2006) Association of the Y402H polymorphism in complement factor H gene and neovascular age-related macular degeneration in Chinese patients. *Invest Ophthalmol.Vis.Sci.* 47 (8):3242-3246.
66. Magnusson KP et al (2006) CFH Y402H confers similar risk of soft drusen and both forms of advanced AMD. *PLoS.Med.* 3 (1):e5.
67. Majewski J et al (2003) Age-related macular degeneration--a genome scan in extended families. *Am.J.Hum.Genet.* 73 (3):540-550.

68. Maller J et al (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat.Genet.* 38 (9):1055-1059.
69. Mandal MN and Ayyagari R (2006) Complement factor H: spatial and temporal expression and localization in the eye. *Invest Ophthalmol.Vis.Sci.* 47 (9):4091-4097.
70. Manuelian T et al (2003) Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J.Clin.Invest* 111 (8):1181-1190.
71. Martin ER et al (2003) Accounting for linkage in family-based tests of association with missing parental genotypes. *Am J Hum Genet* 73 (5):1016-1026.
72. Meyers SM (1994) A twin study on age-related macular degeneration. *Trans.Am.Ophthalmol.Soc.* 92:775-843.
73. Meyers SM, Greene T, and Gutman FA (1995) A twin study of age-related macular degeneration. *Am J Ophthalmol.* 120 (6):757-766.
74. Mori K et al (2007) Association of the HTRA1 gene variant with age-related macular degeneration in the Japanese population. *J.Hum.Genet.* 52 (7):636-641.
75. Mullins RF et al (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.* 14 (7):835-846.
76. Nowak M et al (2004) [The apolipoprotein E polymorphism in age related macular degeneration]. *Klin.Oczna* 106 (3 Suppl):427-428.
77. Okamoto H et al (2006) Complement factor H polymorphisms in Japanese population with age-related macular degeneration. *Mol.Vis.* 12:156-158.
78. Pang CP et al (2000) The apolipoprotein E epsilon4 allele is unlikely to be a major risk factor of age-related macular degeneration in Chinese. *Ophthalmologica* 214 (4):289-291.
79. Pangburn MK, Atkinson MA, and Meri S (1991) Localization of the heparin-binding site on complement factor H. *J.Biol.Chem.* 266 (25):16847-16853.
80. Rivera A 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.* 14 (21):3227-3236.

81. Rosenthal R et al (2006) Expression of bestrophin-1, the product of the VMD2 gene, modulates voltage-dependent Ca²⁺ channels in retinal pigment epithelial cells. *FASEB J.* 20 (1):178-180.
82. Schaffer AA et al (1994) Avoiding recomputation in linkage analysis. *Human Heredity* 44:225-237.
83. Schaumberg DA et al (2006) A prospective assessment of the Y402H variant in complement factor H, genetic variants in C-reactive protein, and risk of age-related macular degeneration. *Invest Ophthalmol. Vis. Sci.* 47 (6):2336-2340.
84. Schaumberg DA et al (2007) A prospective study of 2 major age-related macular degeneration susceptibility alleles and interactions with modifiable risk factors. *Arch. Ophthalmol.* 125 (1):55-62.
85. Schick JH et al (2003) A whole-genome screen of a quantitative trait of age-related maculopathy in sibships from the Beaver Dam Eye Study. *Am. J. Hum. Genet.* 72 (6):1412-1424.
86. Schmidt S et al (2006) Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am. J. Hum. Genet.* 78 (5):852-864.
87. Schmidt S et al (2002) A pooled case-control study of the apolipoprotein E (APOE) gene in age-related maculopathy. *Ophthalmic Genet* 23 (4):209-223.
88. Schmidt S et al (2000) Association of the Apolipoprotein E gene with age-related macular degeneration: Possible effect modification by family history, age, and gender. *Molecular Vision* 31 (6):287-293.
89. Schmidt S et al (2004) Ordered subset linkage analysis supports a susceptibility locus for age-related macular degeneration on chromosome 16p12. *BMC. Genet.* 5 (1):18.
90. Schultz DW et al (2003) Lack of an association of apolipoprotein E gene polymorphisms with familial age-related macular degeneration. *Arch. Ophthalmol.* 121 (5):679-683.
91. Scott WK et al (2007) Independent Effects of Complement Factor H Y402H Polymorphism and Cigarette Smoking on Risk of Age-Related Macular Degeneration. *Ophthalmology.*
92. Seddon JM, Ajani UA, and Mitchell BD (1997) Familial aggregation of age-related maculopathy. *Am J Ophthalmol.* 123 (2):199-206.
93. Seddon JM and Chen CA (2004) The epidemiology of age-related macular degeneration. *Int. Ophthalmol. Clin.* 44 (4):17-39.

94. Seddon JM et al (2003a) Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio. *Arch.Ophthalmol.* 121 (6):785-792.
95. Seddon JM et al (2005) The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch.Ophthalmol.* 123 (3):321-327.
96. Seddon JM et al (2006) CFH gene variant, Y402H, and smoking, body mass index, environmental associations with advanced age-related macular degeneration. *Hum.Hered.* 61 (3):157-165.
97. Seddon JM et al (2003b) A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. *Am.J.Hum.Genet.* 73 (4):780-790.
98. Sepp T et al (2006) Complement factor H variant Y402H is a major risk determinant for geographic atrophy and choroidal neovascularization in smokers and nonsmokers. *Invest Ophthalmol.Vis.Sci.* 47 (2):536-540.
99. Simonelli F et al (2001) Apolipoprotein E polymorphisms in age-related macular degeneration in an Italian population. *Ophthalmic Res.* 33 (6):325-328.
100. Skerka C et al (2007) Defective complement control of Factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol.Immunol.* 44 (13):3398-3406.
101. Smith W et al (2001) Risk factors for age-related macular degeneration: Pooled findings from three continents. *Ophthalmology* 108 (4):697-704.
102. Souied EH et al (1998) The epsilon4 allele of the apolipoprotein E gene as a potential protective factor for exudative age-related macular degeneration. *Am J Ophthalmol.* 125 (3):353-359.
103. Souied EH et al (2005) Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. *Mol.Vis.* 11:1135-1140.
104. Spencer KL et al (2007) Protective Effect of Complement Factor B and Complement Component 2 Variants in Age-related Macular Degeneration. *Hum.Mol.Genet.*
105. Spencer KL et al (2007) Haplotypes Spanning the Complement Factor H Gene Are Protective Against Age-Related Macular Degeneration. *IOVS. In press.*
106. Tanimoto S et al (2006) A polymorphism of LOC387715 gene is associated with age-related macular degeneration in the Japanese population. *Neurosci.Lett.*

107. Tedeschi-Blok N et al (2007) Population-based study of early age-related macular degeneration: role of the complement factor H Y402H polymorphism in bilateral but not unilateral disease. *Ophthalmology* 114 (1):99-103.
108. Thornton J et al (2005) Smoking and age-related macular degeneration: a review of association. *Eye* 19 (9):935-944.
109. Uka J et al (2006) No association of complement factor H gene polymorphism and age-related macular degeneration in the Japanese population. *Retina* 26 (9):985-987.
110. Van LR et al (2004) Cholesterol and age-related macular degeneration: is there a link? *Am.J.Ophthalmol.* 137 (4):750-752.
111. Venables JP et al (2006) Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. *PLoS.Med.* 3 (10):e431.
112. Vingerling JR et al (1995) The prevalence of age-related maculopathy in the Rotterdam Study. *Ophthalmology* 102 (2):205-210.
113. Wang JJ et al (2003) Risk of age-related macular degeneration in eyes with macular drusen or hyperpigmentation: the Blue Mountains Eye Study cohort. *Arch.Ophthalmol.* 121 (5):658-663.
114. Weeks DE et al (2001) Age-related maculopathy: an expanded genome-wide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. *Am J Ophthalmol.* 132 (5):682-692.
115. Weeks DE et al (2004) Age-related maculopathy: a genomewide scan with continued evidence of susceptibility loci within the 1q31, 10q26, and 17q25 regions. *Am.J.Hum.Genet.* 75 (2):174-189.
116. Wong KK et al (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am.J.Hum.Genet.* 80 (1):91-104.
117. Yang Z et al (2006) A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science* 314 (5801):992-993.
118. Yanoff M (2004) *Ophthalmology*. 2nd ed. Mosby: St. Louis, MO.
119. Yates JR et al (2007) Complement C3 Variant and the Risk of Age-Related Macular Degeneration. *N.Engl.J.Med.*
120. Yoshida A et al (2000) Familial cases with age-related macular degeneration. *Jpn.J.Ophthalmol.* 44 (3):290-295.
121. Yoshida T et al (2007) HTRA1 promoter polymorphism predisposes Japanese to age-related macular degeneration. *Mol.Vis.* 13:545-548.

122. Yu J et al (2007) Biochemical Analysis of a Common Human Polymorphism Associated with Age-Related Macular Degeneration. *Biochemistry* 46 (28):8451-8461.
123. Zarepari S et al (2005a) Strong association of the Y402H variant in complement factor H at 1q32 with susceptibility to age-related macular degeneration. *Am J Hum Genet* 77 (1):149-153.
124. Zarepari S et al (2005b) Toll-like receptor 4 variant D299G is associated with susceptibility to age-related macular degeneration. *Hum Mol. Genet* 14 (11):1449-1455.
125. Zarepari S et al (2004) Association of apolipoprotein E alleles with susceptibility to age-related macular degeneration in a large cohort from a single center. *Invest Ophthalmol. Vis. Sci.* 45 (5):1306-1310.
125. Zipfel PF et al (2006) Complement and diseases: defective alternative pathway control results in kidney and eye diseases. *Mol. Immunol.* 43 (1-2):97-106.