MODELING MACULAR DEGENERATION USING QUANTITATIVE PHENOTYPES

By

JOSHUA DAVID HOFFMAN

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

May, 2015

Nashville, Tennessee

Approved:

Tricia Thornton-Wells, Ph.D.

Jonathan L. Haines, Ph.D.

Milam A. Brantley Jr., M.D., Ph.D.

Chun Li, Ph.D.

David C. Samuels, Ph.D.

DEDICATION

To my wonderful wife Cassie, without your endless support and love, none of this would have been possible.

In loving memory of Leonard Greenfield, a wonderful scientist and father-in-law. You will be in our hearts always.

ACKNOWLEDGEMENTS

The body of research presented in this dissertation was supported by grants 1F31AG044089 and 5T32GM080178 (to Joshua Hoffman), AG01972 (to William K. Scott), AG019085, EY012118, and EY023164 (to Jonathan Haines).

Thank you to my dissertation committee at Case Western Reserve University and the Vanderbilt University Medical Center for their support and guidance of my dissertation research: Jonathan L. Haines (advisor), Milam Brantley Jr. (co-advisor), Chun Li (former chair), Tricia Thornton-Wells (chair), and David Samuels (director of graduate studies). I would like to thank Mark Van Grinsven and Clarisa Sanchez at the Radboud University Medical Centre, Netherlands, for their image analysis support.

Thank you to all the current members of the Haines lab: Jessica Cooke Bailey, Mariusz Butkiewicz, Yeunjoo Song, and Rob Igo. Thank you to all the former members of the Haines lab: Nathalie Schnetz-Boutaud, Ping Mayo, Olivia Veatch, Melissa Allen, Brian Yaspan, Anna Cummings, Laura D'aoust, and Mary Davis. Each of you made for a wonderful working environment throughout my studies.

I would like to thank the study participants for providing us with this research opportunity. Thank you to the AMD and Amish ascertainment and clinical teams for collecting the subject related data and paving the way for our analyses. I would also like to thank the administrative staff in the CHGR who made the day-to-day possible.

Thank you to my brother Jonathan Hoffman, my father Richard Hoffman, and my mother Barbara Wolfsdorf for their unending love and support. I would also like to thank my step-father Jack Wolfsdorf, a true inspiration for my career in research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGDB	Anabaptist Genealogy Database
AMD	Age-related Macular Degeneration
anti-VEGF	anti-Vascular Endothelial Growth Factor
AOE	Age of Exam
AREDS	Age-related Eye Disease Study
ARMS2	Age-related Maculopathy Susceptibility Locus 2
BPEI	Bascom Palmer Eye Institute
C2	Complement Component 2
C3	Complement Component 3
CAM	Cell Adhesion Molecules
CAMP	Collaborative Aging and Memory Project
CATT	Age-related Macular Degeneration Trials
CD226	Cluster of Differentiation 226
CFB	Complement component B
CFH	Complement Factor H
CFHR1	Complement Factor Related 1
CFHR1	Complement Factor Related 3
CFI	Complement Factor I
CWRU	Case Western Reserve University
dbGAP	Database of Genotypes and Phenotypes
DHA	Omega-3 long-chain Polyunsaturated Fatty Acids
DUEC	Duke University Eye Center
EPA	Eicosapentanoic acid
EVS	Exome Variation Server
FDA	Food and Drug Administration
GA	Geographic Atrophy
GATK	Genome Analysis Toolkit
GERP	Genomic Evolutionary Rate Profiling
HIHG	Hussmann Institute for Human Genomics
HLOD	Heterogeneity Log-of-odds
HTRA1	Serine Protease Inhibitor 1
indel	Insertion-deletion
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linkage Disequilibrium
LOD	Log-of-the-odds
MQLS	Modified Quasi Likelihood Score Statistic
NFASC	Neurofascin
NRXN1	Neurexin
OCT	Optical-coherence Tomography
RPE	Retinal Pigment Epithelial Layer
SAGA	Swiss Anabaptist Genealogical Association

SCR	Short Consensus Repeat
VEGAS	Versatile Gene-based Association Study
VEI	Vanderbilt Eye Institute

Chapter I

INTRODUCTION

Clinical Features of Age-Related Macular Degeneration

Age-related macular degeneration (AMD) is a debilitating disorder that is responsible for progressive loss of central vision. As the name implies, AMD affects the central portion of the retina, known as the macula. The macula is topologically located 3 mm temporal to the optic disc and has a diameter of roughly 6mm. The macular region is responsible for having the highest visual acuity. AMD is a phenotypically heterogeneous disorder and involves disruption of many aspects of normal retinal anatomy and physiology. These disruptions can be characterized into non-neovascular AMD and choroidal-neovascular AMD. An example of severe AMD as experienced by subjects with the disorder is depicted in Figure 1.1. The known cellular components and layers that play a significant functional role in AMD are the retinal pigment epithelial (RPE) layer, the Bruch's membrane, and the choriocapillaris (Figure 1.2). The RPE layer plays a critical supportive role in normal retinal physiology. The RPE is responsible for regeneration of visual pigments and maintenance of the Bruch's membrane and the overlying rods and cones through transportation of fluids and waste between the choriocapillaris and photoreceptor layer. The Bruch's membrane is an intracellular matrix that acts as a scaffold for the RPE and choriocapillaris, working to regulate cell survival within these two layers[1]. The choriocapillaris, which is a layer of the choroid, is a network of capillaries that acts as the vascular system for the outer retina.

Figure 1.1: Symptoms of AMD Associated Vision Loss

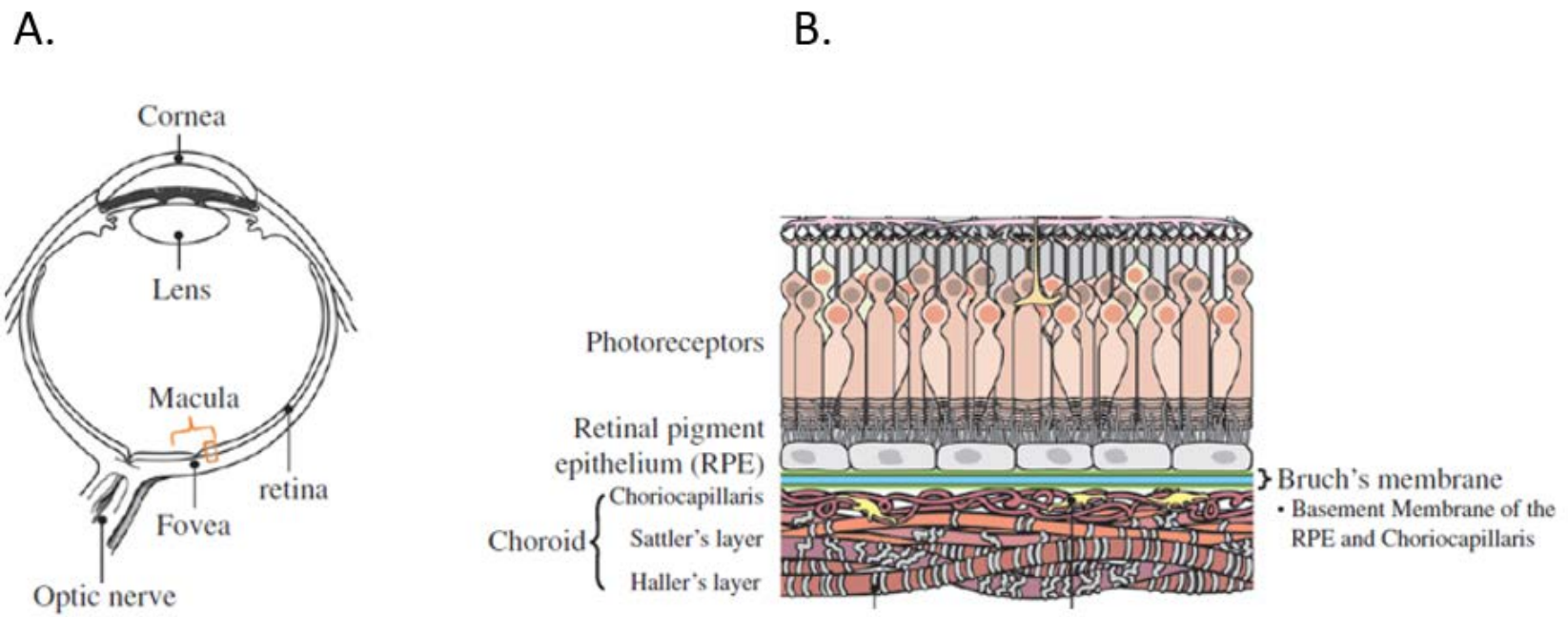
In the left hand image is an example of a person normal eyesight. The right hand image represents an individual with severe AMD associated vision loss. This is a common symptom of late stage wet-AMD

Normal sight

Sight with AMD

Figure 1.2: Major Anatomical Structures of the Eye

A) Diagram of a transverse section of the eye taken through the optic nerve. B) Highlights the major retinal cellular structures that undergo AMD associated physiological changes. Figure is adapted from Campagne et al. 2014[2].



Early stages of Dry AMD

Of the factors that present with AMD, drusen is one the major clinical hallmarks and AMD is rarely diagnosed in their absence[3]. Drusen are extracellular deposits that form between the retinal pigment epithelial (RPE) layer and the Bruch's membrane. Drusen show high variation in morphology and number across individuals and eyes. Drusen morphology can be characterized into hard and soft[4]. Hard drusen are typically less than 63 um in diameter and are observed as small yellow nodules with discrete borders. Soft drusen are commonly greater than 125um in size and take on a pale yellow to greyish appearance and unlike hard drusen, soft drusen tend to have indistinct borders. Subjects may present with small to intermediate drusen throughout their lives and not progress past this stage, while others may form complexes of hard and soft drusen that continually expand over time. It is also important to make the distinction that individuals may present with small drusen and not be considered AMD case subjects.

The biochemical composition of drusen is complex and host a range of polysaccharides, glycosaminoglycans, and lipid components. A major proportion of drusen composition is esterified and non-esterified cholesterol. Protein profiling has estimated 129 different proteins present within hard and soft drusen with some of the common proteins shared between AMD and non-AMD donor eyes being *apolipoprotein B* and *E (APOE), vitronectin, complement factor H, TIMP3*,and complement components *C3,C5,C8,* and *C9*[4-6]. Gene expression analysis of drusen associated mRNA has shown enrichment for *ApoE, vitronectin, TIMP3,* and complement components *C3, C5,* and *C9*[3, 6].

The understanding of drusen formation is ongoing, with many theories existing on how drusen biogenesis may occur. One of the major theories is that drusen formation is a result of inflammation response activation of the complement cascade through immune related signals from the RPE as many of the proteins identified within drusen

are upregulated during immune response[3]. It has also been suggested that drusen formation may be a result of an oxidative stress response[3, 7, 8]. Oxidative modification of extracellular waste produced by the RPE may lead to anchoring of protein components to the Bruch's membrane causing an expansion of drusen over time[3].

Late Stages of AMD

Severe dry AMD is characterized by the presence of geographic atrophy (GA). GA is the result of RPE layer atrophy or cell death and is typically progressive in manner. Due to the supportive nature of the RPE cells, death of this layer is a large factor in secondary loss of photoreceptors and the choriocapillaris leading to deterioration of visual acuity. GA can be uni-focal or multi focal in nature. In unifocal GA, atrophy is localized to a single feature that then spreads outward overtime. As GA progresses it tends to take on a horseshoe appearance, typically surrounding the fovea and eventually this pattern forms a ring later in progression. In multi-focal GA several areas of atrophy present themselves within the macular and spread over time (Figure 1.3 B). In some instances, these multiple areas of atrophy may coalesce into a single atrophic area. Both types of atrophy can later spread to include the foveal region leading to central vision loss[9, 10]. GA progression in this manner may take up to years to develop. Different rates of geographic atrophy progression have been reported, with Holz et. al. showing a median geographic atrophy progression rate of 1.52mm^2/year and Sunnes et. al. reporting 2.2 mm^2/year[11, 12]. Although dry AMD makes up to 20% of all AMD patients that have visual acuity less than 20/200, most of these patients do not have severe vision loss[10].

The more common form of severe vision loss is wet AMD. Wet AMD is thought to arise through a multifactorial process. As part of this process it is hypothesized that an immune response resulting from damage or stress may cause the production of pro-

angiogenic factors such as vascular endothelial growth factor *(*VEGFA*)* by the RPE, leading to choroidal neovascularization. These new blood vessels stemming from the choriocapillaris which are thin and weak, break through a compromised Bruch's membrane and begin to leak blood and fluid into the space separating the Bruch's membrane and the RPE layer[13-15]. This buildup of fluid can cause a blurring of vision at first, and potentially RPE layer detachment. Unlike geographic atrophy, onset of vision loss is a rapid process and can result in blindness of the affected eye within months of first signs and symptoms[1]. Figure 1.3 depicts examples of the clinical hallmarks of AMD as described above.

It is important to note that although AMD may be characterized into different stages and distinct phenotypes, they are all thought to be part of one disease process. Central to this, we find that both wet and dry AMD may co-occur in the same eye, and the underlying risk factors and tissue types affected appear to overlap. This is highlighted by the degenerative changes occurring in the Bruch's membrane and RPE layer during the early to intermediate stages of the disease that then may lead to GA or choroidal neovascularization. The observation that drusen are a risk factor for both severe dry and wet AMD provides further evidence of the interplay between these 2 phenotypes. Research into the potential pathophysiological drivers for the different AMD phenotypes is ongoing.

Figure 1.3 A-D: Examples of AMD Severity
A) Intermediate AMD depicted with soft drusen centered on the fovea and surrounded by small hard drusen. B) Multifocal GA observed as multiple distinct areas of atrophy C) Unifocal GA observed as a flower pattern spreading outwards from the fovea. D) Wet AMD observed as pooling of fluid and blood within the retina.

Epidemiological Studies of AMD

Prevalence estimates for AMD have been carried out for different racial/ethnic populations and varying geographic regions throughout the world. A common finding throughout AMD epidemiological studies is that this disease is relatively rare in individuals under the age of 55, but that prevalence increases dramatically with increasing age groups. Recent reports indicate that there are currently over 1.5 million individuals with AMD in the United States[16]. In a meta-analysis carried out by "The Eye Disease Prevalence Research Group", it was determined that the combined prevalence estimates for both wet and dry AMD in the United States are 1.52% and 16.39% for ages 70-74 and 80, respectively[16].

Variation in prevalence of AMD has been identified between populations of different ancestral backgrounds[16-18]. In a study examining the prevalence of AMD in the Multi-ethnic Study of Atherosclerosis (MESA), Klein et al. observed individuals of European ancestry as having the highest overall prevalence of AMD in the United States when compared to individuals of African American, Chinese, and Hispanic descent (Table 1.1; $p < 0.001$)[16]. When AMD was subdivided into wet AMD and dry AMD, the authors observed a significant increase in frequency of wet AMD in subjects of Chinese descent versus European descent (Table 1.1)[16]. The authors also observed a lower prevalence of wet AMD in African-American versus Caucasian populations, although the presence of large soft drusen along with retinal pigmentary changes seem to occur at similar time points in both racial groups[16].

Table 1.1: Population Specific Prevalence of AMD

For each racial/ethinc group used from the MESA cohort, prevalence estimates of the different AMD subtypes were calculated per separate age and gender groups. Taken from (Klein et al. 2006)[16].

	Early AMD									Late AMD								
	Whites		*Blacks*		*Hispanics*		*Chinese*			*Whites*		*Blacks*		*Hispanics*		*Chinese*		
Gender and Age* (yrs)	*No. at Risk*	%	*No. at Risk*	%	*No. at Risk*	%	*No. at Risk*	%	P *Value*†	*No. at Risk*	%	*No. at Risk*	%	*No. at Risk*	%	*No. at Risk*	%	P *Value*†
Men																		
45–54	296	1.7	198	0.5	185	2.7	99	0.0	0.22	296	0.0	198	0.0	185	0.0	99	0.0	
55–64	316	3.2	211	1.9	191	2.6	96	3.1	0.84	317	0.3	211	0.0	191	0.0	97	1.0	0.35
65–74	341	6.5	230	1.3	165	3.6	100	5.0	0.02	343	0.6	231	0.4	166	0.0	101	1.0	0.67
75–84	163	12.9	84	2.4	70	11.4	47	14.9	0.02	169	3.6	85	0.0	73	1.4	50	6.0	0.11
Total	1116	5.2	723	1.4	611	3.9	342	4.4	<0.001	1125	0.8	725	0.1	615	0.2	347	1.4	0.02
P value‡				<0.001									0.03					
Women																		
45–54	327	1.8	252	0.4	210	1.4	102	1.0	0.47	327	0.0	252	0.0	210	0.0	102	0.0	CNE
55–64	361	2.5	253	2.8	183	2.2	103	1.0	0.86	361	0.0	253	0.0	185	1.1	103	0.0	0.06
65–74	326	4.6	252	2.8	187	5.3	101	5.0	0.50	327	0.0	253	0.0	187	0.0	101	0.0	CNE
75–84	169	13.6	103	8.7	83	12.0	43	7.0	0.55	175	2.3	107	2.8	83	0.0	46	4.3	0.32
Total	1183	4.5	860	2.8	663	4.1	349	2.9	0.18	1190	0.3	865	0.3	665	0.3	352	0.6	0.90
P value‡				0.24									0.90					
Both genders																		
45–54	623	1.8	450	0.4	395	2.0	201	0.5	0.09	623	0.0	450	0.0	395	0.0	201	0.0	CNE
55–64	677	2.8	464	2.4	374	2.4	199	2.0	0.96	678	0.1	464	0.0	376	0.5	200	0.5	0.25
65–74	667	5.5	482	2.1	352	4.5	201	5.0	0.02	670	0.3	484	0.2	353	0.0	202	0.5	0.61
75–84	332	13.3	187	5.9	153	11.8	90	11.1	0.06	344	2.9	192	1.6	156	0.6	96	5.2	0.11
Total	2299	4.8	1583	2.1	1274	4.0	691	3.6	<0.001	2315	0.6	1590	0.3	1280	0.2	699	1.0	0.06
P value‡				<0.001									0.08					

CNE = cannot estimate.
*Age at visit 1.
†From Fisher exact test: tests null hypothesis that there is not a racial difference for the given age/gender category (across all ages [gender] for the total row).
‡Cochran–Mantel–Haenszel test controlling for age and gender (where appropriate).

Environmental and Clinical Risk Factors Associated with AMD

Age is currently the biggest risk factor for developing AMD. The disease is the least common in individuals under the age of 55 and the most common in individuals over the age of 75[8,9,19-22]. Studies examining the risk associated with gender have identified conflicting evidence of gender associated risk of AMD[16, 22, 23]. Some studies have shown a significantly higher prevalence of AMD among females in Caucasian populations, especially in older ages groups (age at exam > 60)[16, 22, 23]. When examining subjects aged between 55-84, Smith et al. showed a significant increase in risk for females (OR = 1.15)[24]. The second major environmental risk factor that has been associated with AMD consistently is smoking, with a relative risk between 2-4 for individuals that smoke compared to those that have never smoked[25-30]. Cross sectional and longitudinal studies carried out by the Blue Mountain Study group with respect to cardiovascular risk factors for AMD have shown that high-density lipoprotein and cholesterol may be inversely related to AMD risk, and that presence of cardiovascular disease such as stroke, myocardial infarction, and angina may be a risk factor[31]. The effects of hypertension on a person's risk for developing AMD have been contradictory, but BMI has been identified as a risk factor for AMD[32, 33]. Other studies examining the role of sunlight/UV exposure in AMD risk have not yielded conclusive results[34-36].

Treating AMD

To examine the role that vitamin supplementation may play in slowing AMD progression, a clinical trial carried out by the Age Related Eye Disease Study (AREDS) research group looked specifically at the effect of high doses of anti-oxidants and zinc on

AMD progression and visual acuity[37]. The results showed that subjects who were put on a high-antioxidant plus zinc nutritional supplementation plan had a significantly reduced risk of progressing to advanced AMD versus the placebo group (OR = 0.72), but had little effect on progressing from early to intermediate stage AMD[37]. AREDS 2, examined the potential protective impact of lutein + zeaxanthin, omega-3 long-chain polyunsaturated fatty acids (DHA) and eicosapentanoic acid (EPA) on AMD progression[38]. The results of this study have recently been completed and showed that changes to the current AREDS formulation significantly altered AMD progression compared to placebo[38]. Results of the AREDS2 trial suggested replacing beta-carotene with lutein and zeaxanthin while the addition of fatty acids had no significant effect[38].

Although the use of vitamins and anti-oxidants has shown some promise in slowing AMD progression, the major tool used in the treatment of AMD has been the use of anti-Vascular Endothelial Growth Factor (anti-*VEGF* antibodies) wet AMD. Two of the major anti-VEGF treatments currently being used in the clinical setting are ranibizumab (Lucentis) and bevacizumab (Avastin) as an off-label drug. Lucentis is a humanized monoclonal antibody fragment that binds to and blocks all forms of *VEGF-A*. In one of the pivotal clinical trials, 95% of individuals receiving intravitreal injections of ranibizumab had lost less than three lines on the standard visual acuity chart for the treated eye after 1 year compared to only 62% of those receiving placebo injections[39] . Avastin, which is also a humanized monoclonal antibody, was originally Food and Drug Administration (FDA) approved for the treatment of colon cancer, but was found to slow progression of wet AMD in subjects receiving this drug. In the “Comparison of Age-related Macular Degeneration Trial” (CATT), which examined the efficacy of both Avastin and Lucentis, both drugs had similar impacts on AMD treatment[41]. One of the newer treatment regimes that has recently been approved by the FDA is aflibercept (EYLEA).

This molecule is a *VEGF* trap, preventing *VEGF* from binding to its receptors. EYLEA has the benefit of increased time between injections compared to Avastin and Lucentis, with dosages occurring once a month in the first 3 months of treatment followed by bimonthly injections thereafter.

Genetic Epidemiology of AMD

Although AMD is a complex, phenotypically heterogeneous disorder, the presence of a genetic role for AMD has been well documented through twin studies and familial aggregation studies. In a familial aggregation study carried out by Seddon et al., first-degree relatives of those affected with late AMD were at an increased odds of risk (OR = 2.4) when compared to families that did not have members with AMD[31]. A twin study carried out by Seddon et. al. in 840 twins estimated the heritability of intermediate AMD to be 0.67 and advanced AMD to be 0.71[42]. The group also reported an overall AMD heritability estimate of 0.46[42]. Similar findings were reported by Hammond et. al. when their study examined 406 twin pairs that produced heritability estimates between 0.45 to 0.81, depending on the severity of the disease[43].

Early Genetic Studies of AMD

One of the first major tools used to dissect the genetic architecture of AMD were genome-wide linkage studies. Genome-wide linkage analysis is carried about by examining the segregation of genetic markers spaced throughout the genome with a given trait or disease within families. Generally, a log-of-the-odds (LOD) score greater than three is considered significant evidence for linkage. One of the earliest linkage screens carried out for AMD was in a single 21 family member pedigree by Klein et al. [44].

Linkage analysis in this family identified a maximum multi-point LOD score of 3.0 in the 1q25-q31 region that segregated in a dominant fashion. Following the work by Klein et al., several other linkage screens were carried out proposing regions 9q31, 10q26, 12q23, 15q21, 16p12, 17q25, and 22q13 as showing evidence of linkage[45-49]. A meta-analysis carried out by Fisher et al. to examine previous linkage peeks that were only moderately associated with the AMD phenotype, identified strong linkage on chromosome 1q and 10q26, and significant evidence of linkage on chromosomes 2p,3p,4q,12q, and 16q[50].

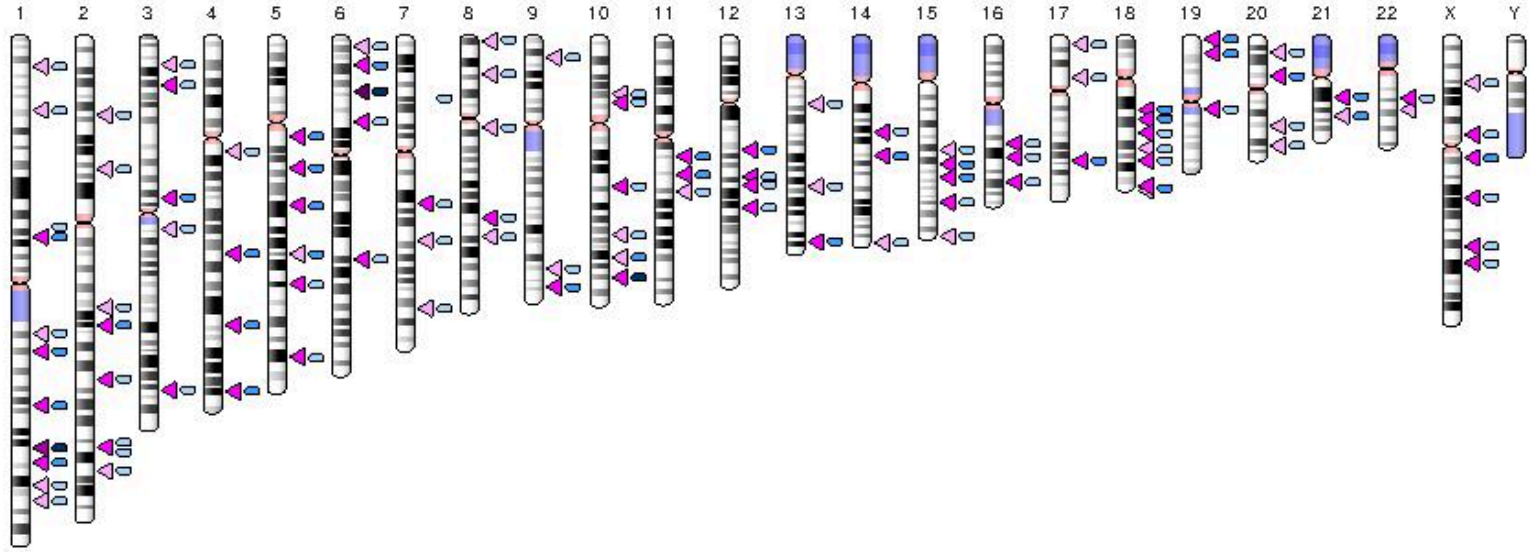
With the combination of success found in linkage screens for AMD and its highly heritable nature for a complex common disease, AMD provided a strong candidate for genome-wide association studies (GWAS). GWAS arose out of the completion of the Human Genome Project in 2000 and the estimation of linkage disequilibrium across the genome by the International HapMap Consortium in 2002[51, 52]. Through these efforts, panels made up of single marker polymorphisms (SNPs) ranging from 100,000 to over 1,000,000 markers were created to interrogate the genome. GWAS allowed for hypothesis-free association testing of genetic markers with disease across the entire genome, and thus provided a powerful statistical approach for identifying the contribution of unknown common genetic variation to disease.

The application of GWAS and fine mapping of SNPs to AMD proved successful with several studies publishing association results simultaneously in 2005[50, 53-57]. These studies implicated a non-synonymous variant in a previously reported linkage peak that causes a serine to histidine amino acid change (Y402H) in C*omplement Factor H* (*CFH*)[53-56]. Several variants within the10q26 locus were also shown to be strongly associated with AMD and implicated the region containing *age-related maculopathy susceptibility locus 2* (*ARMS2*) / *Serine Protease Inhibitor 1* (*HTRA1*) as the genes of

interest[50, 57]. Most of the success of GWAS in AMD can be attributed to the large magnitude of effect that these variants exhibit, and the frequency of the risk alleles in the population, a property that is relatively rare in complex common diseases. The *CFH* risk allele was proposed to have an estimated OR of over 2.5 per allele and a risk allele frequency of approximately 0.54 in cases and 0.36 in controls[53-56]. The variants harbored in *ARMS2*/*HTRA1* had an estimated allelic OR of 2.7 and a risk allele frequency of 0.42 in cases and 0.20 in controls[50, 57].

Genome-wide linkage and association analyses have implicated a multitude of variants across the entire genome as risk factors for AMD. An integrative phenotype-genotype search of known association results returned 331 SNPs in or around 225 different genes across all 22 autosomes and the X chromosome, highlighting the potential extent of genetic signals contributing to AMD risk (Figure 1.4).

Figure 1.4: Genetically Associated AMD Loci
Ideogram showing variants genetically associated with AMD. Arrows point to specific genes. *



*http://www.ncbi.nlm.nih.gov/gap/phegeni#GenomeView; accessed January, 2015

Complement Factor H

CFH is located on chromosome 1 within a cluster of genes that encode the regulatory components for the activation of C3 and has been one of the most significantly replicated genes to be associated with AMD to date. *CFH* is a glycoprotein that plays an inhibitory role within the complement pathway. *CFH* selectively binds and inactivates complement component C3b thereby preventing the continuation of the complement cascade.

The *Y402H* variant rs1061170 in *CFH* was one of the first SNPs found to be associated with AMD [45-49]. The magnitude of the effect of this variant on AMD risk is considered substantial, with Haines et al. reporting odds ratios between 2.45 and 5.57 for having one or two copies of the risk allele respectively[55]. Klein et al. and Edwards et al. also presented similar finding with reported odds ratios around 2.7 per risk allele[53, 56, 58].

Further work carried out at the *CFH* locus suggests that there may be other disease-associated variants present throughout the gene. In a study carried out by Li et al. examining 84 polymorphisms in and around the *CFH* locus, 20 polymorphisms showed stronger association with AMD than the *Y402H* variant [43]. They also observed that no single polymorphism could explain the role of *CFH* in AMD risk, but instead that there was the presence of four common haplotypes along with multiple rare haplotypes with two of the common haplotypes being associated with AMD risk and two being protective. The pathophysiological role that *CFH* plays in AMD may be explained in part due to the observation of complement proteins including *CFH* within drusen, as it is thought that local inflammation and activation of the complement pathway may in some part be responsible for AMD pathogenesis. In work by Clark et al., *CFH* was found to be

the major regulator of the complement cascade at the Bruch's membrane[51]. Further work showed the highly associated *Y402H* variant resulted in reduced binding of *CFH* to the Bruch's membrane, thus potentially leading to increased activation of the complement cascade and damage to this layer[59, 60].

Other Complement Pathway Associated Genes

CFH has not been the only complement pathway gene identified as a contributor to AMD risk. In a case-control study that interrogated single nucleotide polymorphisms across the *Complement component 3* (*C3*) and *Complement component 5* (*C5*) loci, Yates et al. showed that a polymorphism in *C3* to be significantly associated with AMD with an odds ratio of 2.6 for individuals that were homozygous for the risk allele[61]. This polymorphism has been confirmed by an independent study carried out by Maller et al. in two independent Caucasian populations[62]. Like *CFH*, *C3* localizes to drusen. *Complement component B* (*CFB*) and *Complement component C 2* (*C2*), located in the major histocompatibility complex III region, are two other genes that have been implicated in AMD. The minor allele of in C2/CFB has been found to be associated with reduced risk (OR = 0.57) [63]**.** In a case-control haplotype analysis carried out by Gold et al., one common risk haplotype (OR = 1.32) and two protective haplotypes (OR = 0.45 for H7 and 0.36 for H10) were identified at the *CFB*/*C2* locus[63]. Real-time PCR performed by this group also showed the presence of CFB and C2 in ocular drusen and within the Bruch's membrane[63]. The protective role of one of the *CFB/C2* haplotypes may be linked to reduced hemolytic activity within the complement cascade, and as such a reduction in inflammation response. In addition to the *CFH* protective haplotypes, copy number variation analysis identified a *complement factor related* (*CFHR1/CFHR3*) deletion that confers protection against AMD[64]. Functional analysis of the *CFHR3*

deletion identified this locus as an inhibitor of *C3* convertase activity, thus playing a role in the complement pathway[65].

Age-Related Maculopathy Susceptibility 2 and Serine Protease Inhibitor 1

ARMS2 and *HTRA1* located on chromosome 10 is another major locus implicated in AMD pathogenesis. There has been much debate as to the causal locus inferring AMD risk, as the 2 associated variants, rs11200638 in *HTRA1*'s promoter and rs10490924 in *ARMS2* (A69S) are located only 6.6 kb away from another, and LD between these variants is high (D'=0.97,r^2 = 0.93)[66, 67]. As such, these genes are commonly referred to as a single locus (*ARMS2*/*HTRA1*). In an in-depth case-control analysis of the 10q26 locus, Yang et al. reported ORs as high as 6.5 for the homozygous genotype at SNP rs11200638[68]. At the time, the authors estimated that this SNP contributes a population-attributable risk close to 50% for AMD, and when taken in conjunction with the *CFH* variant Y402H, population attributable risk may reach as high as 71%[68].

Functional analysis of the *HTRA1* gene suggests that it may be involved in the regulation of extracellular matrix proteoglycan degradation. Overexpression of *HTRA1* may lead to a reduction in the integrity of the Bruch's membrane allowing for the invasion of choriocapillaris across the extracellular matrix, a feature that is present in the wet form of AMD. This potential physiological role is supported statistically by genetic association analyses that have shown differences in the magnitude of effect when adjusting late AMD subjects for the presence of GA or wet AMD[69, 70]. In these analyses, the authors observed that *CFH* attributed higher risk to subjects with bilateral GA compared to subjects with bilateral wet AMD. The authors described variants in *ARMS2*/*HTRA1* as conferring higher risk to bilateral wet AMD. Although these

magnitude of effect differences exist, risk variants in these genes still show high statistical significance with both phenotypes[69, 70]. Expression analysis of the HTRA1 gene in the presence of the rs11200638 variant has led to conflicting results, with some studies suggesting an association with increased mRNA expression levels while other studies have not been able to replicate these results[71-74, 74, 75].

Little is known about the biological role of *ARMS2*. Evolutionarily it is thought to have arisen recently within the primate specific lineage[76]. Fritsche et al. identified an insertion-deletion (indel) polymorphism within *ARMS2* conferring risk in the presence of at least one copy of the deletion (OR = 2.9)[76]. It was shown that harboring two copies of the deletion resulted in undetectable levels of *ARMS2* in subject donor eyes. Fritsche et. al. also observed that ARMS2 localized to the mitochondria and that *ARMS2* may act through mitochondrial mediated pathways[76]. Contradictory to this study, Wang et al. reported that *ARMS2* localized to the cytosol and not the mitochondria of ARPE-19 and COS7 cells. In a separate study by Wang et al., the A69S SNP was observed to be in high LD with the indel and did not correlate with *ARMS2* mRNA expression[77].

Contribution of Rare Variants to AMD

The application of high-throughput sequencing has allowed investigators the opportunity to shift from common variant to rare-variant analyses for the study of AMD to identify the missing contributors of AMD genetic risk[78, 79]. Recent studies have identified a number of rare variants in known AMD associated genes that confer risk for AMD. Ven et al. identified a highly penetrant rare mutation in complement factor I (*CFI*) that associated with AMD risk with an OR of 22 in a European-based sample[80]. Functional analysis showed reduced expression of *CFI* and less efficient degradation of *C3b* in carriers of this variant. In a whole-genome sequencing study examining 2,200 Icelandic subjects, a rare non-synonymous *C3* variant was identified conferring an OR of 3.45 and

follow up functional analysis showed reduced binding of C3b to CFH[81]. The authors were able to replicate this statistical association in an independent European sample where an even larger magnitude of effect was observed (OR = 4.22).

In a recent analysis in which the authors sequenced all genes that harbor common variants reported to be associated with AMD in over 1,600 cases and 750 controls, the authors identified a significant enrichment of variants in cases compared to controls based on gene burden tests[82]. Interrogation of individual variants within these genes showed significant associations with rare-variants in *CFI*, *C3*, and *C9*[82]. In a functional analysis of the rare *C3* variant, the authors showed reduced inactivation of *C3* by *CFH* and *CFI*[82].

Common Variants of Low Effect

Although the bulk of common variants of modest to large affect have been identified, a recent meta-analysis was completed to identify the remaining common genetic contributors to AMD through a large meta-analysis[69]. As one of the most significant meta-analysis to date in terms of subject sample size, the authors replicated 12 previously identified AMD loci and reported 7 new loci of minor to modest effect in over 17,000 cases and 60,000 controls[69]. The authors predicted that these 19 common variants contribute 15% to 65% of the total genetic variation attributed to AMD risk based on prevalence estimates of 1% to 30%. The seven new loci by themselves only predict to contribute 1% of the genetic variation explained for AMD[69]. The effect sizes of the replicated and novel loci are shown in Table 1.2. Examination of these 19 loci in a cumulative genetic risk score analysis showed that these 19 loci can modestly predict AMD disease status with an area under the curve (AUC) of 0.74.

Table 1.2: List of Common Variants Reaching Genome-Wide Significance in an AMD Meta-Analysis
Figure taken from Fritsche et. al, 2013[69]

SNP	Risk allele	Chr.	Position	Nearby genes	EAF	Discovery		Follow-up		Combined	
						P	OR	*P*	OR	*P*	OR (95% CI)
Loci previously reported at $P < 5 \times 10^{-8}$											
rs10490924	T	10	124.2 Mb	*ARMS2-HTRA1*	0.30	4×10^{-353}	2.71	2.8×10^{-190}	2.88	4×10^{-540}	2.76 (2.72–2.80)
rs10737680	A	1	196.7 Mb	*CFH*	0.64	1×10^{-283}	2.40	2.7×10^{-152}	2.50	1×10^{-434}	2.43 (2.39–2.47)
rs429608	G	6	31.9 Mb	*C2-CFB*	0.86	2×10^{-54}	1.67	2.4×10^{-37}	1.89	4×10^{-89}	1.74 (1.68–1.79)
rs2230199	C	19	6.7 Mb	*C3*	0.20	2×10^{-26}	1.46	3.4×10^{-17}	1.37	1×10^{-41}	1.42 (1.37–1.47)
rs5749482	G	22	33.1 Mb	*TIMP3*	0.74	6×10^{-13}	1.25	9.7×10^{-17}	1.45	2×10^{-26}	1.31 (1.26–1.36)
rs4420638	A	19	45.4 Mb	*APOE*	0.83	3×10^{-15}	1.34	4.2×10^{-7}	1.25	2×10^{-20}	1.30 (1.24–1.36)
rs1864163	G	16	57.0 Mb	*CETP*	0.76	8×10^{-13}	1.25	8.7×10^{-5}	1.17	7×10^{-16}	1.22 (1.17–1.27)
rs943080	T	6	43.8 Mb	*VEGFA*	0.51	4×10^{-12}	1.18	1.6×10^{-5}	1.12	9×10^{-16}	1.15 (1.12–1.18)
rs13278062	T	8	23.1 Mb	*TNFRSF10A*	0.48	7×10^{-10}	1.17	6.4×10^{-7}	1.14	3×10^{-15}	1.15 (1.12–1.19)
rs920915	C	15	58.7 Mb	*LIPC*	0.48	2×10^{-9}	1.14	0.004	1.10	3×10^{-11}	1.13 (1.09–1.17)
rs4698775	G	4	110.6 Mb	*CFI*	0.31	2×10^{-10}	1.16	0.025	1.08	7×10^{-11}	1.14 (1.10–1.17)
rs3812111	T	6	116.4 Mb	*COL10A1*	0.64	7×10^{-8}	1.13	0.022	1.06	2×10^{-8}	1.10 (1.07–1.14)
Loci reaching $P < 5 \times 10^{-8}$ for the first time											
rs13081855	T	3	99.5 Mb	*COL8A1-FILIP1L*	0.10	4×10^{-11}	1.28	6.0×10^{-4}	1.17	4×10^{-13}	1.23 (1.17–1.29)
rs3130783	A	6	30.8 Mb	*IER3-DDR1*	0.79	1×10^{-6}	1.15	3.5×10^{-6}	1.16	2×10^{-11}	1.16 (1.11–1.20)
rs8135665	T	22	38.5 Mb	*SLC16A8*	0.21	8×10^{-8}	1.16	5.6×10^{-5}	1.13	2×10^{-11}	1.15 (1.11–1.19)
rs334353	T	9	101.9 Mb	*TGFBR1*	0.73	9×10^{-7}	1.13	6.7×10^{-6}	1.13	3×10^{-11}	1.13 (1.10–1.17)
rs8017304	A	14	68.8 Mb	*RAD51B*	0.61	9×10^{-7}	1.11	2.1×10^{-5}	1.11	9×10^{-11}	1.11 (1.08–1.14)
rs6795735	T	3	64.7 Mb	*ADAMTS9*	0.46	9×10^{-8}	1.13	0.0066	1.07	5×10^{-9}	1.10 (1.07–1.14)
rs9542236	C	13	31.8 Mb	*B3GALTL*	0.44	2×10^{-6}	1.12	0.0018	1.08	2×10^{-8}	1.10 (1.07–1.14)

Studying Genetically and Environmentally Isolated Populations

The study of a complex disease such as AMD introduces a number of genetic, phenotypic, and environmental variables that must be dissected to find contributing risk factors to disease. A majority of studies that examine genetics of AMD take large, population based approached to identify attributable genetic risk loci. One measure of reducing this heterogeneity is through the study of environmentally and genetically isolated founder populations. With these types of populations, we observe a more homogenous environmental and genetic background, potentially increasing our power to detect genetic associations within complex diseases. The Amish communities of Indiana and Ohio are one example of an isolated study population that we study extensively.

The Amish represent a genetically isolated founder population that emigrated from Western Europe in two waves beginning in the early 1700's. During the first wave, members of the Swiss Anabaptist Community immigrated to Pennsylvania. In a second wave of immigration occurring in the early 1800's, members of the Pennsylvania Amish community moved to Holmes County in Ohio. In addition to the local establishment of this new community in middle Ohio, another group of immigrants arriving from Western Europe established communities in Elkhart LaGrange and Adams County in Indiana[83, 84].These immigration patterns created population bottlenecks, causing a reduction in genetic variation within these new communities. The Amish live a strict religious and cultural lifestyle, typically marrying within the faith and generating large families. This characteristic, in addition to relatively recent founder event that created this community, creates a more genetically and environmentally homogenous group as compared to the general population.

Another advantage of using the Amish population for genetic studies is their extensive genealogical record keeping. Records on members of the Amish community are continuously updated in the Anabaptist Geological Database (AGDB) and the Swiss Anabaptist Genealogical Association (SAGA)[85, 86]. This record keeping allows all members of the community to be connected into a single pedigree. This extensive pedigree information proves valuable for measuring degrees of relationship and querying extended families.

The use of this Midwestern Amish population has already proved successful in elucidating genetic loci in complex disease as applied to studies of late-onset Alzheimer's disease (LOAD) and Parkinson Disease (PD)[87-89]. Through linkage and association analysis, Cummings et. al. identified *CTNNA2* as a potential risk locus for LOAD and additionally identified 3 other loci with log odds (LOD) score over 3[89]. It was also shown that the major risk allele in APOE was significant in this population, showing that this population can be informative for application to other European ancestry based populations[89].

Conclusion

AMD is a complex and debilitating disorder with many environmental and genetic components contributing to the overall risk of developing the disease. As a disease of aging, we can expect that the number of new cases to appear in the coming years to increase at an alarming rate as a result of technological advances in medicine increasing average life expectancy. Continued understanding of the implications that both known and unknown genetics risk factors play in AMD is necessary. Even with the success that genetic and epidemiological studies have brought to understanding AMD risk, the

functional roles of these variants have proven difficult to discern and generally are conflicting. Much of this complexity may be explained by the seemingly high overlap between the AMD phenotypes and physiological pathways that are involved in the manifestation of disease. To dissect this interplay, it is necessary to perform phenotype specific analyses of AMD. With drusen playing such a pivotal role in AMD diagnosis and risk for disease progression, we explore the potential impact of known genetic variation on this phenotype (Chapter II & III).

Most, if not all, the common variants that contribute at least modest effects to AMD have been identified. It is now also of importance to away at the unknown genetic component of AMD through rare variant analysis. In chapter IV, we explore the impact of rare genetic variation on AMD using an isolated founder population that has shown reduced cumulative genetic risk when put in context of the major known AMD genetic risk loci.

Although there has been some success in the development of new treatments for AMD, most if not, all are targeted towards the advanced stages. Understanding the contribution of known genetic variation to the pathological changes observed in AMD and identifying the missing genetic component are paramount. The elucidation of these processes will allow us to continue the development and the refining of how AMD risk is assessed and screened, especially in the early diagnosis of this disease.

Chapter II

DESCRIPTION OF STUDY POPULATIONS AND QUALITY CONTROL PROCEDURES

Introduction

Fundus photographs used in our analyses were selected from two separate datasets: one coming from the Age-related Eye Disease Study (AREDS) and a second one from the combined dataset from Case Western Reserve University (CWRU) and the Hussman Institute for Human Genomics (HIHG). The AREDS dataset was used in our study as it was available within the “Database of Genotypes and Phenotypes” (dbGAP). Through the contribution of AREDS to dbGAP, a wealth of deep phenotype data across thousands of individuals is publicly available. In total, 650 variables ranging from demographic data such as race, sex, and age, to clinical measures such as blood pressure, cholesterol and history of current prescription drug use, as well as history of other eye or non-eye related disorders were made available. Ascertainment of the CWRU/HIHG dataset has been ongoing for over the past 10 years and is continually expanding. Subjects were recruited from Durham North, Carolina; Nashville, Tennesee; and Naples and Palm Beach, Florida. CWRU/HIHG is a clinic based sample and was originally focused on collecting multiplex families for the study of AMD. Shifting study designs has led to the collection of subjects for case-control and longitudinal treatment and progression studies. A further description of the CWRU/HIHG dataset can be found in chapter III. The AREDS dataset will be discussed in further detail concerning sample collection, grading criteria, and inclusion/exclusion criteria below.

AREDS Study Design

AREDS was designed to identify the impact of high doses of antioxidants and zinc on cataracts and AMD. AREDS was divided into three separate phases and spanned a total of 16 years. The major goal of phase one was to design the study and enroll and screen subjects for eligibility and took place from March 1990 to October of 1992.

Phase two was the clinical trial portion of the study, which aimed to define the role of antioxidants and zinc on AMD and cataract progression. Phase two took place from November 1992 through September 2001 in which time subjects deemed eligible in phase one, were observed for up to seven years. During phase two, study participants were assigned to a treatment regimen consisting of either antioxidants, zinc, a combination of antioxidants and zinc, or a placebo. Subjects that had a diagnosis of less than intermediate AMD were not assigned to the zinc group due to the risks associated with zinc intake. Description of the AREDS grading scale will be discussed further in the grading classification section of this chapter. As part of the phase two study design, subjects were followed-up every six months with a clinical exam. Macular photographs were taken at the enrollment and randomization visit, which was set as the baseline, then two years from baseline. Following the two-year visit, photographs were taken annually. If visual acuity dropped by more than 10 letters during the time from the baseline visit to any of the follow-up visits, a non-annual photograph was required. In total, subjects' median follow-up time for phase two was 6.5 years. The major outcome variables that were measured during phase two were the morphological changes associated with drusen, geographic atrophy, and choroidal neovascularization, along with changes in visual acuity and function.

The objective of phase three was to understand the natural history of AMD progression by extending the clinical trial for another five years. By increasing the observation time of subjects, the AREDS investigators' hypothesis was that more subjects would progress from the control group to cases, and controls that did not progress would make a valuable control cohort due to their non-progressing to slowly progressing nature. Another major objective of the phase three trial was to generate new grading scales to better predict disease progression. As an extension of phase two, subjects continued to be seen annually for ophthalmic examination although subjects were not required to continue their previous treatment regimen.

Ascertainment and clinical exams of subjects for AREDS took place at 11 different clinical centers throughout the U.S. These centers included the Eye Center at Memorial in New York, Associated Retinal Consultants in Pennsylvania and Michigan, Devers Eye Institute in Oregon, Emory University in Georgia, Massachusetts Eye and Ear Infirmary, National Eye Institute Clinical Center in Maryland, University of Pittsburgh Eye and Ear Institute in Pennsylvania, Ingalls Memorial Hospital in Illinois, Johns Hopkins in Maryland, Elman Retina Group in Pennsylvania and Maryland, and the University of Wisconsin. Although routine clinical follow-up and fundus photography was carried out at these 11 institutions, grading of fundus photographs was completed at a centralized reading center at the University of Wisconsin.

Use of Color Fundus Photography for Traditional Grading in AMD

Use of color fundus photography has been the standard and most effective imaging tool employed by retinal specialists in the diagnosis and manual grading of

intermediate AMD to date[90]. Many grading systems have been presented that use fundus photography for diagnosis, but one of the most common, and the one employed by the AREDS study, is the Wisconsin Age-related Maculopathy Grading System[91]. In this system, a grid consisting of three circles measuring 500um, 1500um, and 3000um in radius, and four radial lines that are concentric with the center of the macula, is placed on a plastic sheet and superimposed on the fundus image (Figure 2.1). In addition to the placement of this grid on the center of the macula, the retinal specialist uses a reference set of open circles corresponding to different sized area measurements. A trained grader or retinal specialist uses these circles to sum the total surface area covered by the feature to then assign a grade (Figure 2.1).

On the Wisconsin Scale, drusen are measured and categorized into small, intermediate, and large. Small drusen are those that have a diameter less than 63um, medium sized drusen have a diameter in the range of 63um to 125um, and large drusen are those that are larger than 125um. Intermediate AMD is characterized by the presence of a substantial number of medium sized drusen or by the presence of one large druse located within 3000um of the center of the macular region (Figure 2.2 A-D).

Figure 2.1: Age-related Maculopathy grading grid with reference circles.

Within the grid the three circles with 4 radial lines originating from the center represent radii of 500 um, 1500um, and 3000um. On the right are the typical set of open circles of varying sizes for estimating the size of AMD related abnormalities including GA and drusen. Circle C-0 represents 63um in diameter, C-1 equals 125um in diameter, c-2 equals 250 um, I-1 equals 175um, I-2 equals 350um, O-1 equals 332 um, and O-2 equals 644 um. Adopted from Klein et. al[91].

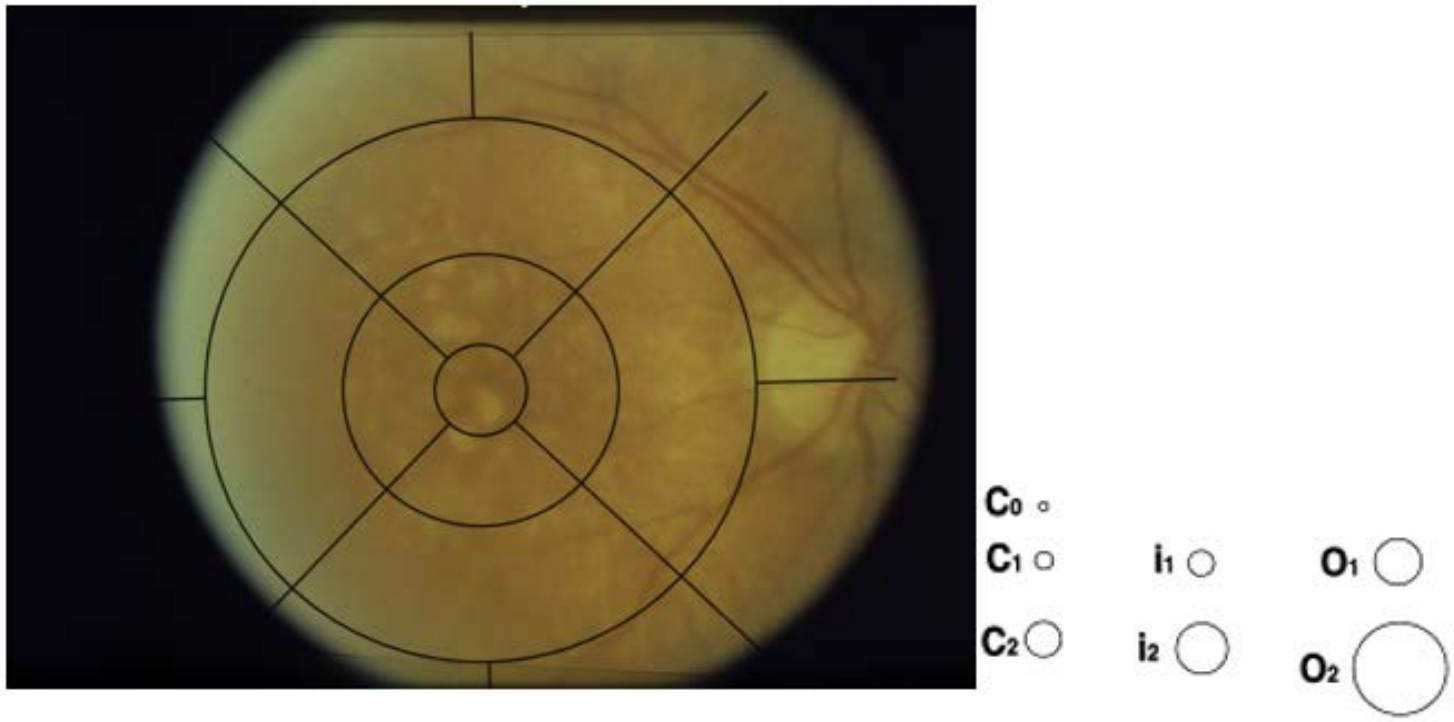
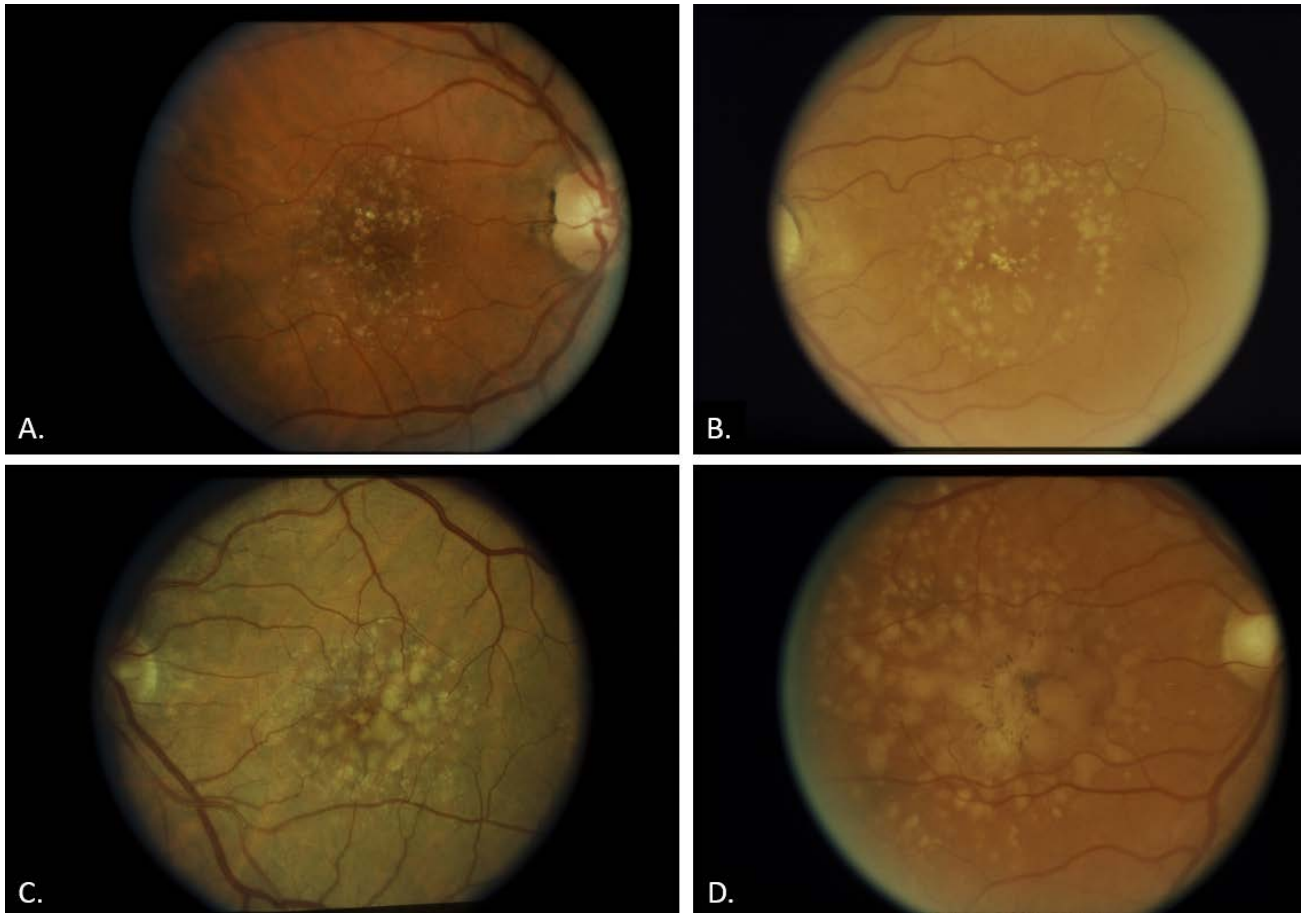


Figure 2.2: A-D: Drusen of Varying Severity and Type

In this figure we observe varying sizes and number of drusen that present themselves within the retina. The images are to highlight the tedious and difficult process that is required to accurately quantify drusen. Figure A. We observe an abundance of small to intermediate hard druse. Figure B-C. We observe small, intermediate, and large hard and soft drusen. Figure D. We observe large soft confluent drusen.



Cohort Grade Classification and Eligibility Requirements

AREDS

Categorization of AREDS subjects into AMD severity groups for the clinical trial was based upon a one through four-grade classification scheme. A grade of one corresponded to both eyes of a subject having at most small non-extensive drusen and no pigment abnormalities. Subjects in this grade must also have had a corrected visual acuity score greater than 74 letters. A grade of two corresponded to at least one of the subject's eye having either extensive small drusen, multiple intermediate drusen, or pigment abnormalities that have been known to be associated with AMD. These subjects must not have had large drusen or advanced AMD in either eye or a visual acuity score less than 73 letters.

Subjects with a grade of three are broken down into two (A and B) subcategories. Subjects with a grade of 3A have at least one eye with one or more of the following; one or more large drusen, intermediate drusen covering at least that of circle I-2 in the presence of soft drusen, intermediate drusen covering the area at least of circle O-2 if soft indistinct drusen are not present. These subjects also may have had geographic atrophy if it was not observed within the center of the macula. In addition, subjects with a grade of three must not have had advanced AMD in either eye or a visual acuity score of 73 or less letters if associated with AMD. Subjects in subcategory 3B met the requirements for 3A in one eye and the fellow eye has a visual acuity score less than 73 letters that is not associated with AMD.

Subjects with a grade of four could be broken down into two subcategories, 4A and 4B. Subjects with a grade of 4A are those that showed the presence of advanced

AMD in one eye but not the other, and a visual acuity score that was or was not less than 73 letters or less. Subjects with a grade of 4B were those that had a visual acuity score less than 74 letters and no presence of advanced AMD. The fellow eye for this subject must also have had a visual acuity score greater than 74 and no advanced AMD. A brief summary of the AREDS classification scheme is given in Table 2.1.

To be eligible for the phase two clinical trial, the following visual acuity requirements must have been met per subject. Subjects given a grade of one or two must have had a visual acuity greater than 74 letters in each eye, subjects assigned a AMD grade of three must have had a visual acuity of 74 or more in at least one eye with large drusen or geographic atrophy not included in the central macula region. Subjects given a grade of four in one eye must have had a visual acuity score greater than 74 letters in the fellow eye if advanced AMD was not present. Subjects given a grade of four in both eyes must have had a visual acuity of 74 or more in one eye and 73 or less in the other eye.

Table 2.1: AREDS AMD Classification Scheme.

Table taken from The Age-related Eye Disease Study Group et. al 1999[92]

AMD Category	First Eye[a]			Second Eye
	Drusen Size[b]	Drusen Area[b]	Pigment Abnormalities (PA)[c]	
Category 1	None or small (<63μm)	<125μm diameter circle (≈ 5–15 small drusen)	None	Same as first eye
Category 2	Small (<63 μm) or	⩾125 μm diameter circle	Absent or present, but geographic atrophy (GA) absent	Same as first eye or category 1
	Intermediate (⩾63, <125 μm) None required, if pigment abnormalities present	At least one druse		
Category 3a	Intermediate (⩾63 μm, <125 μm)	⩾360 μm diameter circle, if soft indistinct drusen are present (≈20 intermediate drusen)	Absent or present but central GA[b] absent	Same as first eye or category 1 or 2
	or	⩾656 μm diameter circle (about 1/5 disc area), if soft indistinct drusen are absent (≈65 intermediate drusen)		
	Large (⩾125 μm) None required, if non-central GA[b] present	At least one druse		
Category 3b	First eye same as category 3a			VA <20/32 not due to AMD; or uniocular disqualifying disorder is present
Category 4a	First eye same as category 1, 2, or 3a			Advanced AMD[d]
Category 4b	First eye same as category 1, 2, or 3a			VA <20/32 due to AMD, but advanced AMD[d] not present

[a] Must have visual acuity ⩾20/32, no advanced AMD, and no disqualifying lesions.

[b] Drusen and geographic atrophy are assessed within 2 disc diameters of the center of the macula.

[c] Pigment abnormalities (increased pigmentation or depigmentation) within 1 disc diameter of the center of the macula.

[d] Geographic atrophy (GA) involving center of macula or signs of choroidal neovascularization (presence beneath the retinal pigment epithelium or sensory retina of fluid, blood, or fibrovascular or fibrous tissue).

The age requirements for subjects to be included in phase two were that subjects with an AMD category of one or two needed to be of age 60 through 80 at the time of qualifying visit, and subjects with a grade of three or four needed to range from 55 through 80 at the qualifying visit. Subjects that were over the age of 78 at the time of registration for phase one were age eligible for the phase two qualifying visit regardless of AMD severity grade.

In addition to the age and visual acuity requirements, subjects were also screened for the presence of a history of retinal abnormalities other than AMD. Abnormalities warranting exclusion included central serous choroidopathy, optic atrophy, surface wrinkling retinopathy, retinal pigmentary changes that are not specific to AMD, macular hole, retinal vein occlusion, or general ocular diseases that could potentially impact the assessment of AMD or cataracts. In addition, subjects that showed the presence of diabetic retinopathy were excluded. To reduce the enrichment of subject eyes with glaucoma, subjects that had an intraocular pressure of greater than 26mm Hg or a medical history eluding to the potential presence of glaucoma were given a visual field exam to look for the presence of glaucomatous visual field defects. Subject eyes presenting with any of these glaucoma related abnormalities were excluded.

Surgical based exclusions included subjects with a history of laser photocoagulation for peripapillary CNV or for AMD in the eye receiving treatment. Subjects who received cataract surgery within 6 months of the phase 2 qualifying visit were also excluded for that eye.

CWRU/HIHG

Information about the inclusion criteria and ascertainment for the CRWU/ HIHG cohort can be found in chapter III. The exclusion criterion for this cohort included any

case that has had a history of retinal surgery, laser photocoagulation, glaucoma, diabetic retinopathy, or any other eye-related disease.

The grading scale used by the CWRU and HIHG for the diagnosis of AMD is a derivation of the classification scheme used by AREDS and is described in Table 2.2**.** The major modification of the AREDS scale is the addition of a fifth category to differentiate advanced AMD into geographic atrophy (grade four) and choroidal neovascularization (grade five). As this is a step scale, subjects that present with both geographic atrophy and neovascularization are categorized as grade five.

Table 2.2: CWRU and HIHG Grading Scale
Table taken from Spencer et. al 2007[93]

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

Selection of Samples for Drusen Analysis

Information about the selection criteria of subjects from both the AREDS and CWRU cohorts can be found in chapter III.

Image Quality Control

To accurately quantify drusen load, it is necessary to have images of sufficient quality to detect drusen of varying size (small, medium, large) and type (hard and soft). As such, prior to analysis of drusen progression, we employed both computational assessed quality score measures and subjective quality score measures to both the AREDS and CWRU/HIHG datasets. It is important to note that capturing of fundus photos by the AREDS clinical centers was completed in a non-digital format. The cameras employed at these centers were the Zeiss FF series of cameras, which took analogue fundus photographs using transparency type film. Clinical centers were approved to use either Kodachrome or Ektachrome film with this camera type.

To reduce variability in images over time, centers were required to only use one film type over the course of the entire study, although each center was free to choose which of these films to use. Within the variables given by AREDS, there is no note on which film type was selected at each reading center and thus cannot be assessed as a potential bias in drusen quantification. Subjective information about photographic quality was completed as part of AREDS; this information was not made available to the public on a per image basis. It must also be noted that quality assessment by the AREDS grading center was completed on the native non-digital images and not the digitized images. Thus, although digitization of fundus photographs on a whole may not lead to major artifactual differences between native image and the digitized images, we are aware that we are analyzing copies that are further processed into JPGs by AREDS.

In contrast to the AREDS fundus photographs captured using transparency film, fundus photographs captured as part of the CWRU and HIHG dataset were taken

digitally using either a Zeiss 450 (Carl Zeiss, Jena, Germany) camera at the Vanderbilt Eye Institute, or a Topcon TRC 50IX camera (Topcon 50IA, Tokyo, Japan) at the Bascom Palmer Eye Institute.

As part of the quality control process, we performed a first pass removal of poor quality images based on manual image inspection. Major criteria for image removal included blurriness to a point where small drusen could not be delineated, improper distance of camera from the eye, lateral misalignment of the camera, presence of the grading grid on the image, major variation in contrast across the photograph, images missing the full macular field, and absence of the optic nerve as this is used in alignment of the grading grid. Examples of these major quality issues are presented in Figure 2.3 A-H. Images from the CRWRU/HIHG dataset were screened prior to the onset of this study and were not confounded by these artifacts.

Figure 2.3 A-H: Example of Major Artifacts Present in AREDS Photographs

Figure A. Large color variation across fundus field. Figure B. Lateral misalignment of camera. Figure C. Optic nerve absent from image. Figure D. Grading grid left on image

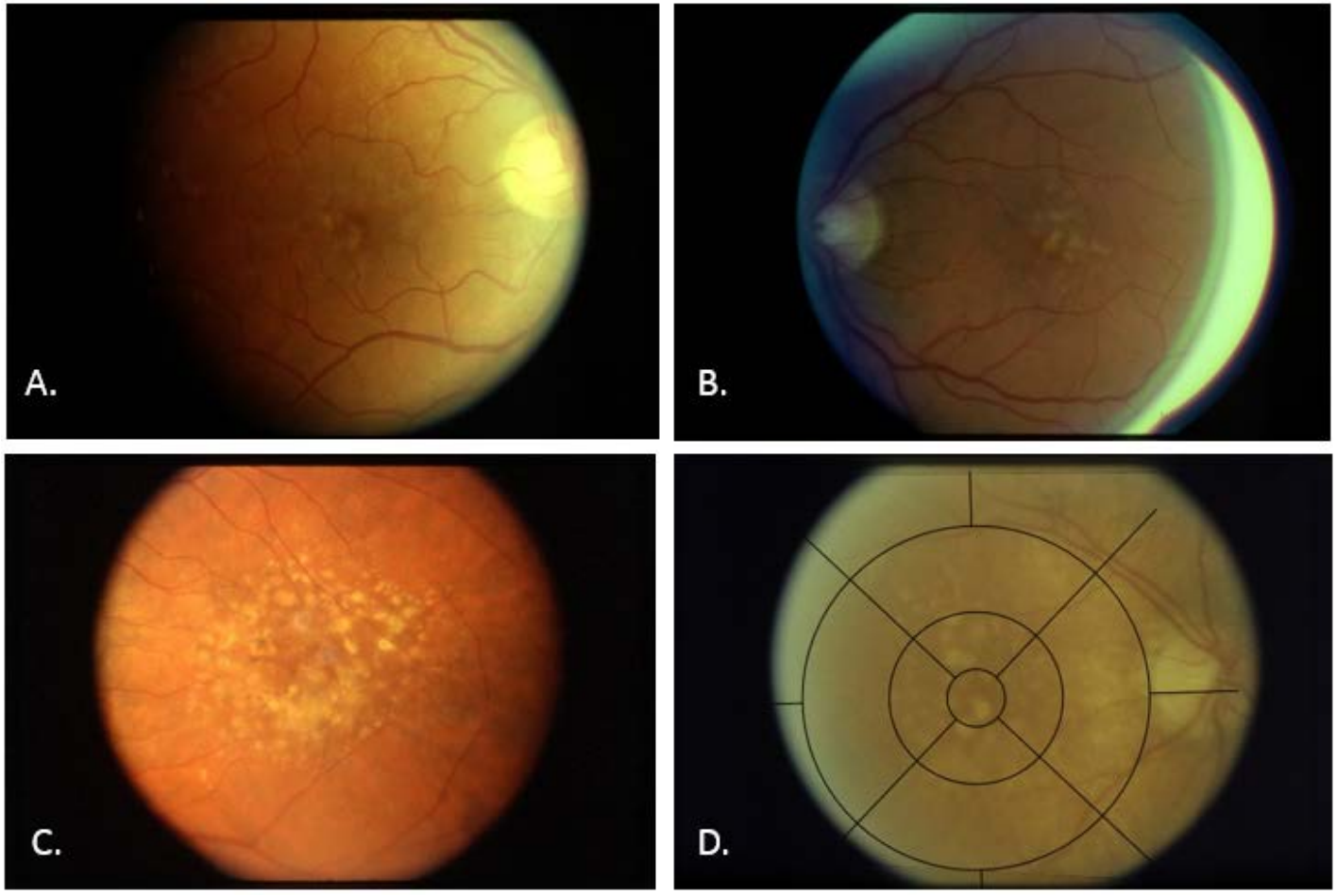
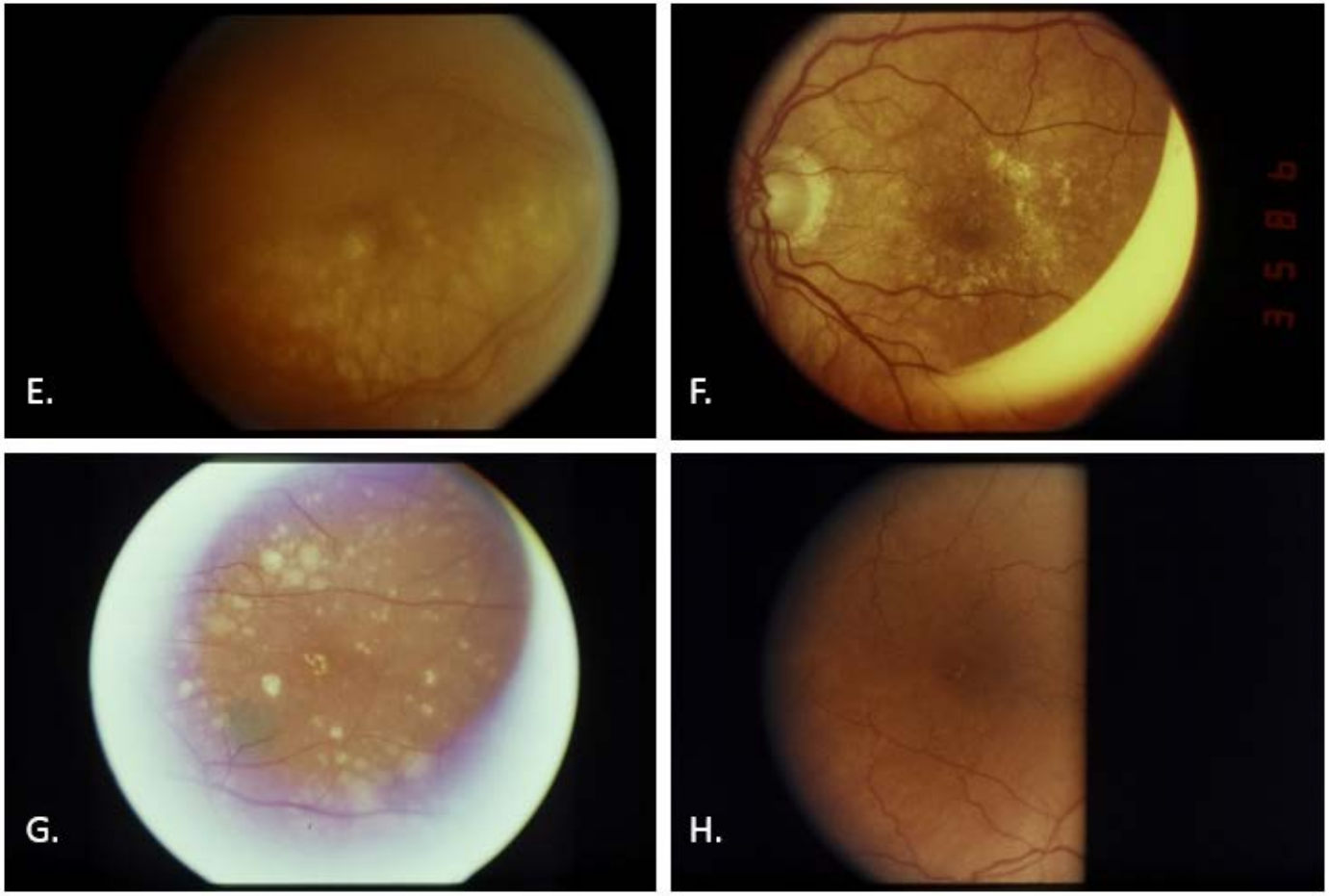


Figure 2.3 A-H: Example of Major Artifacts Present in AREDS Photographs

Figure E. Image too blurry to delineate drusen. Figure F. Camera too close to eye Figure G. Camera too far from eye. Figure H. Cutoff image



After removal of images based on gross artifacts, we performed a computational assessment of image quality using a retinal workstation package developed at Radboud University Medical Center[94]. Quality assessment with this package is calculated as a continuous measure from zero, being very poor, to a score of one being very good. Metrics involved in calculating the quality score include the level of blurriness, contrast in the image, level of brightness, and ability to delineate anatomical features such as the optic nerve and blood vessels surrounding the macula. Results from the quality assessment differed greatly between the CWRU/HIHG dataset and the AREDS dataset. When we examined the distribution of quality scores in the Case/Western HIHG dataset, we see a bimodal distribution with most images being of high quality according to the quality assessment algorithm (Figure 2.4 A). When we examine the quality distribution of the AREDS data we again see a bimodal distribution, although in these data most images fall into the low quality range (Figure 2.4 B).

Visual inspection of images given a low quality score by the algorithm did not generalize well to manual image inspection as we observed many cases of images being given a low quality score in the AREDS dataset but being of good quality when visualized manually. When we group quality scores into categories of less than 0.25 and greater than 0.75, as this is where most of the images fall in the quality distribution, we do not observe a significant difference in quality score versus drusen area based upon a Wilcoxon Rank Sum test (p-value = 0.06;Figure 2.5). This led us to conclude that quality score data given by the retina workstation package was not having an effect on drusen area quantification. Based on these qualitative and quantitative results the quality score did not appear to have value and was not used or investigated further.

Figure 2.4 A-B: Quality Distribution of CWRU/HIHG and AREDS Images
Figure A. CWRU/HIHG Quality Distribution.

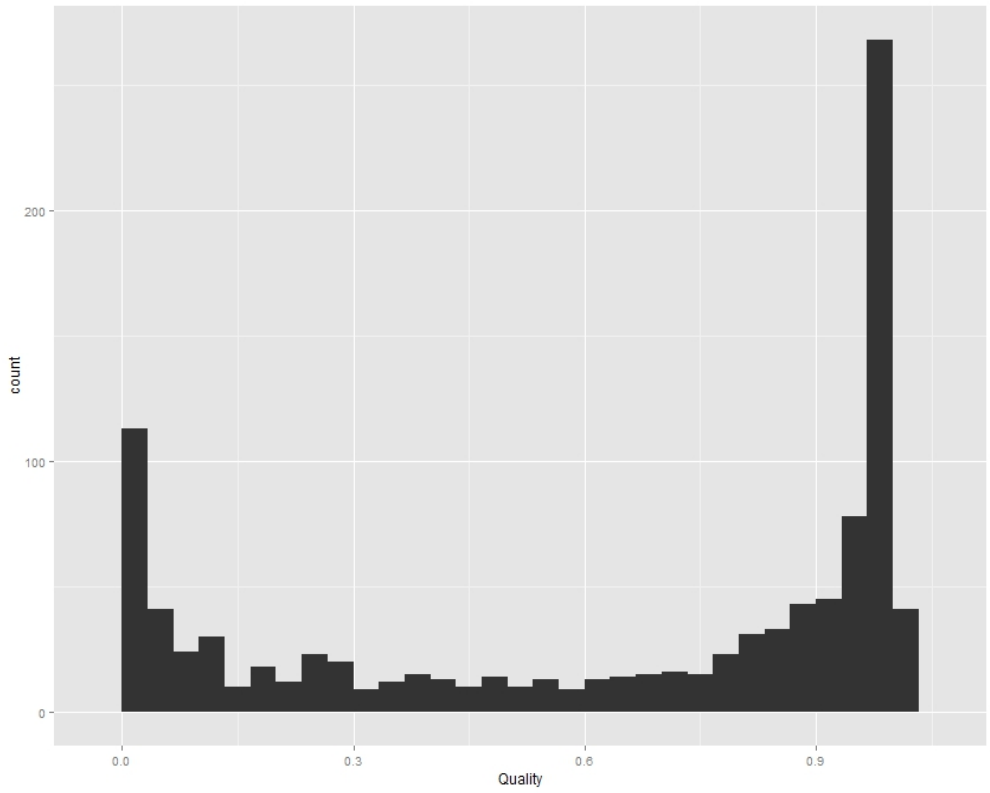


Figure 2.4 A-B: Quality Distribution of CWRU/HIHG and AREDS Images
Figure B. AREDS quality distribution.

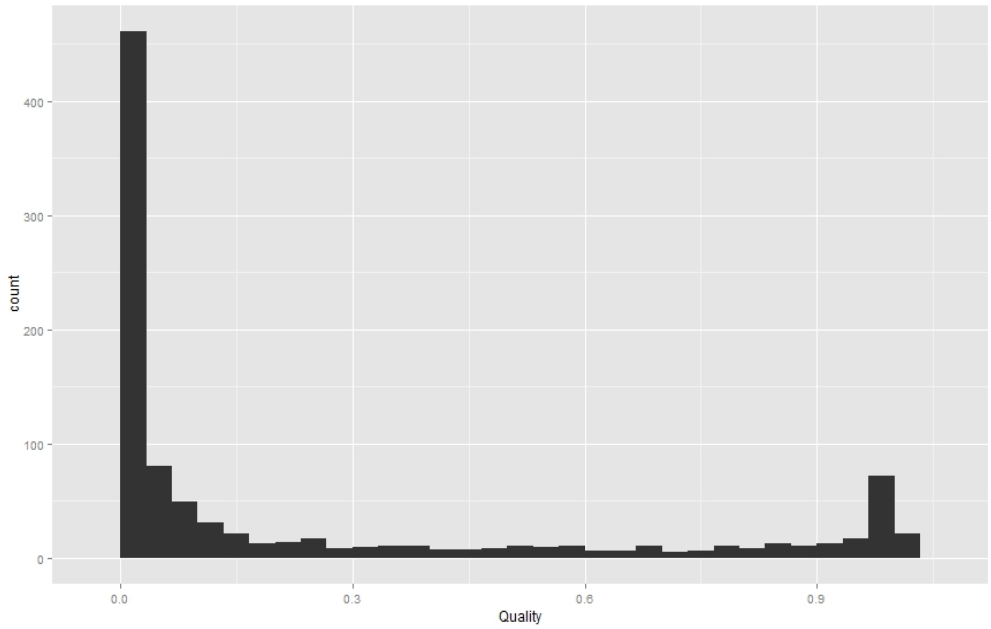
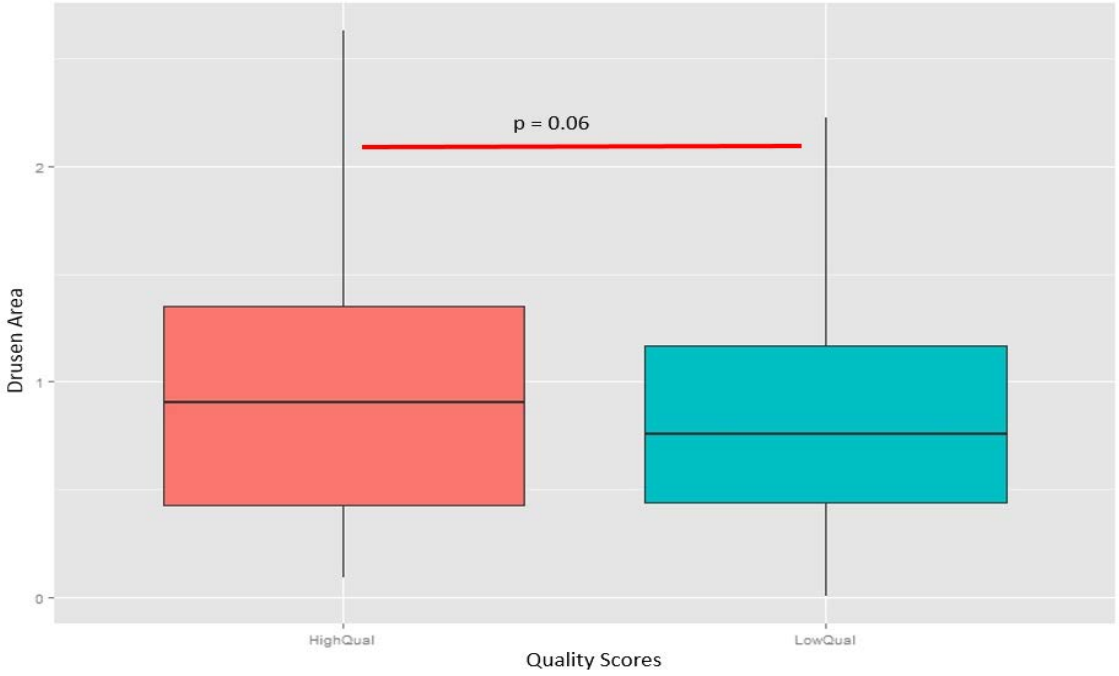


Figure 2.5: Effect of Quality Score on Drusen Calls
Boxplot examining differences in drusen area when stratifying images by low quality (< 0.25) and high quality (> 0.75).

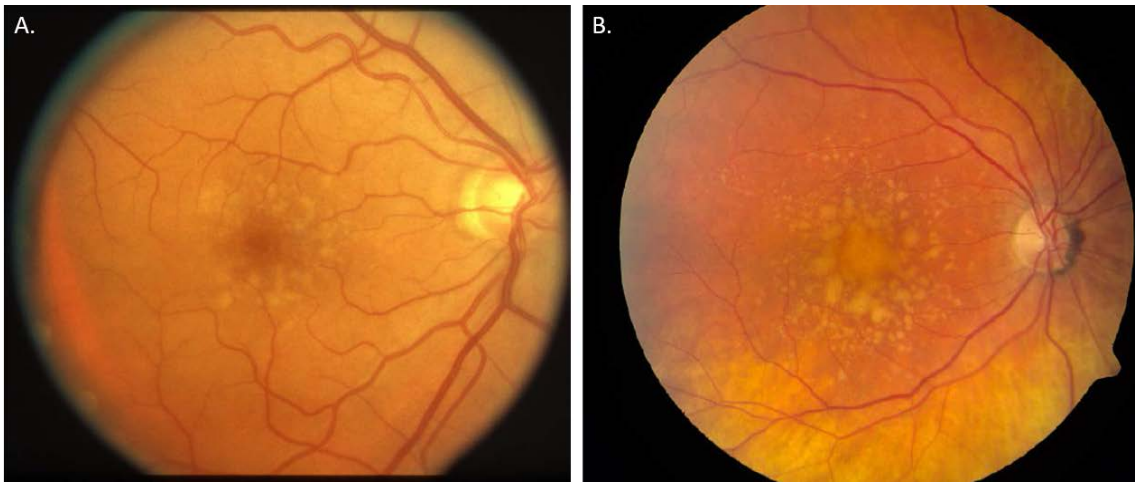


Macular Grading Grid Placement and Sizing

As part of the quantification algorithm, a macular grading grid is automatically placed on color fundus photographs to set the region of interest for quantification. Similar to manual grading, the digital grading grid is placed on the center of the fovea and the outermost grid circle is lined up with the edge of the optic disc. The algorithm standardizes the radial distance as 3,000um to translate the pixels of the digital image into the metric scale. This relative scaling information was important to our quantitative analysis as both the CWRU/HIHG photographs and the AREDS photographs were taken using different magnification thresholds. As is presented, AREDS photographs were 30-degree photos and the CWRU/HIHG photos were 50-degree photos (Figure 2.6 A-B). Both types of photographs were centered on the macula. Although different

magnifications were employed by the different centers, analysis of drusen is limited to areas contained within the grading grid where the fovea is the central reference point. The degree of the photo refers to the magnification scale used in the image. 30-degree photo is magnified 2.5X larger than life while 50-degree photos translate to 1.5X magnification.

Figure 2.6 A-B: Magnification Differences Between 30 and 50 Degree photographs
Figure A. 30 degree fundus photograph as used by AREDS. B. 50 degree fundus photograph as used by CWRU/HIHG



As discussed in chapter III, all images were resized to 650 pixels regardless of photographic magnification. To adjust for these magnification differences within the drusen quantification step, a consensus pixel estimate was determined through anatomical measurements of the distance between the center of the fovea and the edge of the optic disc. A consensus grid radius of 210 pixels was concluded for the AREDS 30-degree photos, and a 150-pixel radius was concluded for the CWRU/HIHG 50-degree photographs. Through visual examination of the digital grading grid placed on

photographs with physical grading grids left attached to AREDS photos, it was determined that our consensus 210 pixel radius was exceedingly concordant. Radial sizing of the CWRU/HIHG macula grading grid was confirmed by a retinal specialist for proper anatomical sizing (M.A.B. at the VEI).

Drusen Identification Threshold

Although the quantification algorithm is an automated process, the user has an option of specifying a sensitivity/specificity threshold to determine at what level a drusen call should be considered a true call. Drusen that are detected by the algorithm are assigned a probability indicating the likelihood that the region contains a true druse. This threshold is a sliding scale running from zero, which is high sensitivity, to one, indicating high specificity . For our analysis, we chose 0.5 as the threshold based on empirical testing by the Diagnostic Imaging and Analysis Group at Radboud University, Nigmegen. This threshold was set to include only drusen that were detected with a high probability.

Examination of CWRU/HIHG Fundus Photograph Replicates

Using the parameters discussed in the previous sections, we set out to examine the impact of varying image contrast and positioning on drusen quantification. As part of the collection of fundus photographs in the CWRU/HIHG dataset, multiple fundus photographs were taken of a subject's eye per visit. Although these images are capturing similar amounts of information, we observe differences in image quality throughout these replicates. Variation included changes in contrast and brightness, slightly different angle of the image with relation to the macula, and subtle variation in image focus. In these data, we examined 156 replicates that fit the criteria mentioned above. Within these replicates we observe very high replicate correlation, with the major

source of variation being due to large changes in focal point and blurriness (Spearman's correlation = 0.97; Figure 2.7)

Assessing Categorical Versus Quantitative Measures of Drusen Area

Following removal of poor quality images and after selection of subjects as described in chapter III, we are left with 973 images across 276 subjects in the AREDS dataset. Introduced, as tabulated data within AREDS, was a categorized drusen area variable (dbGAP variable accession number: phv00054045.v1.p1). As part of this variable, drusen area was measured by a retinal specialist and graded on a scale from 0-7 at each photographed visit. Assignment of drusen area was completed using the methods described in the "Use of Color Fundus Photography for Traditional Grading of AMD" section. Matching the quantification results with the retinal specialist grades, we observe a steady increase when comparing continuous results versus graded results (Figure 2.8).

Figure 2.7: Results of CWRU/HIHG Replication Analysis.
In this figure, we examine the effect of image variation on quantification results. Major sources of variation included image blurriness, focal point, contrast and brightness.

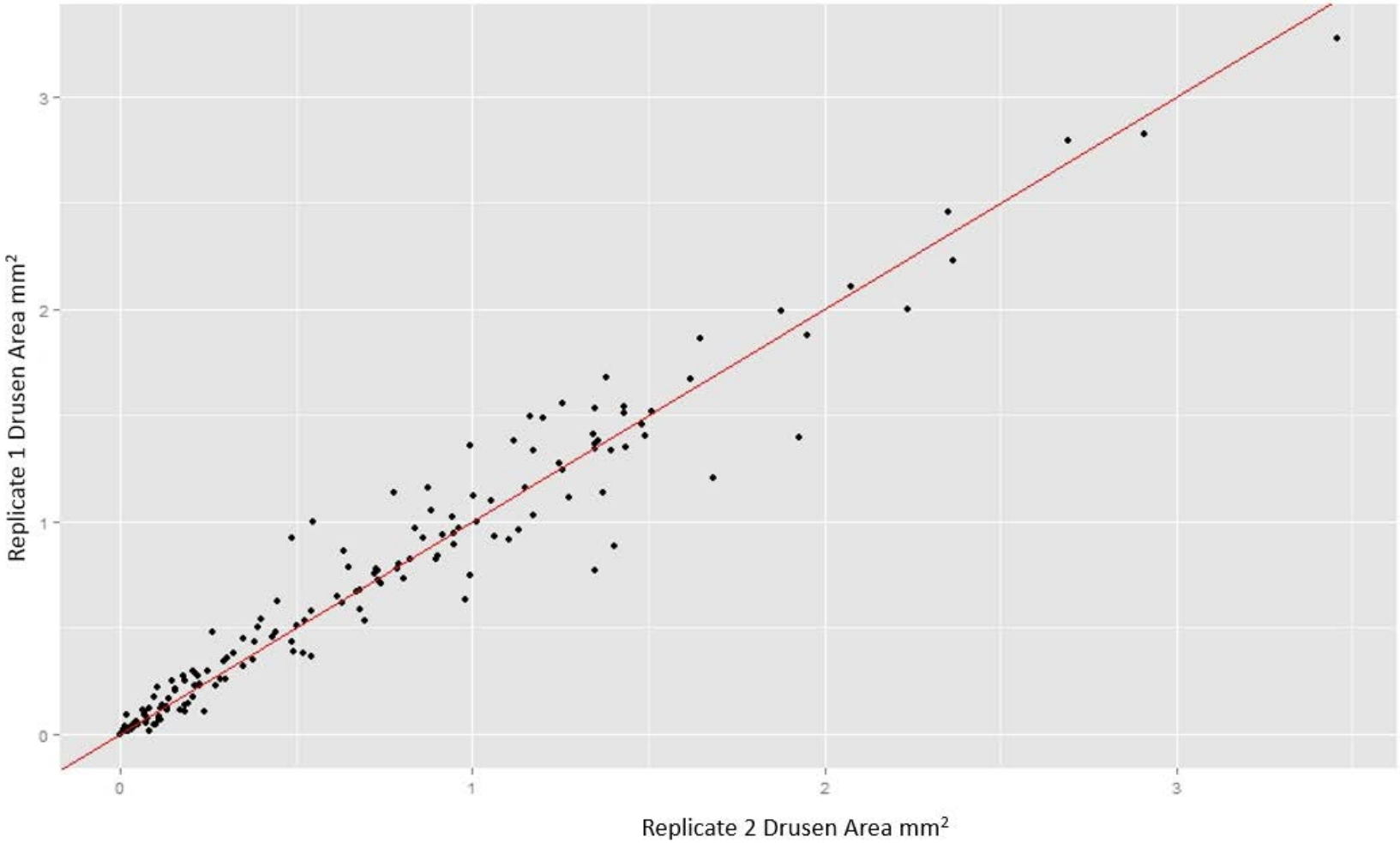
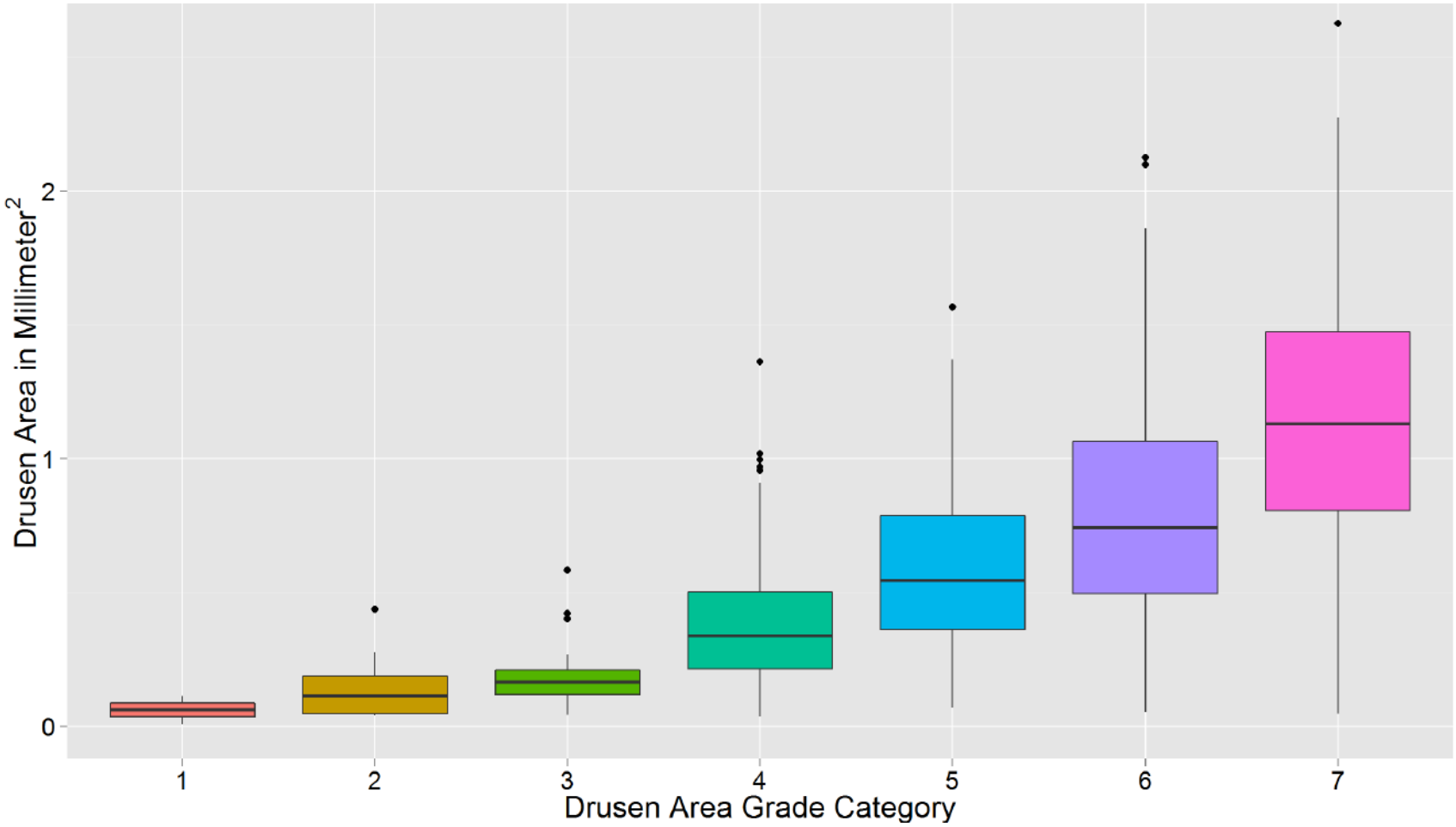


Figure 2.8: Quantified Versus Categorized Drusen Area
Examination of quality controlled image quantification results versus manual assigned area grades.



Discussion

Grading of drusen is a long tedious process and the ability to accurately quantify this feature manually is difficult due to the complexity in size, shape, and type of drusen that present themselves within the macula (Figure 2.2 A-D). Application of manual grading to large scale studies can be a daunting task and require a great number of resources in the form of human hours devoted by a trained retinal specialist. Inter-grader variability also poses an issue due to the sometimes-subjective nature of drusen boundaries especially in the face of retinal pigmentary abnormalities and numerous small and intermediate drusen. High-throughput automated quantification of drusen provide the ability to overcome the issues present with manual grading and to apply novel methods to an imaging modality that has been use to collect data for decades. Although refinement of these methods is ongoing, we chose to use a previously developed algorithm that fully automates the drusen quantification process[95]. The algorithm from which we present our results takes advantage of color fundus photographs to extract drusenoid features and calculate total surface area covered by drusen within the age-related eye disease study maculopathy grading grid.

The necessity of having high quality images for quantification is paramount for any automated method as this can greatly increase the power of detecting and calling true drusen within fundus photographs. Although the AREDS images are described as needing to be of sufficient quality to assess the drusen severity spectrum from small to large drusen, it quickly became apparent that major quality issues arose upon visual inspection of these images. After removal of fundus photographs with gross artifacts, we are left with 973 images in the AREDS dataset that are of a subjectively sufficient quality for drusen calling. We computationally assessed image quality with a software package

on the retinal workstation, although the accuracy of the results were not indicative of true image quality and did not reflect inaccuracy in drusen calling.

Analysis of the replicate data shows strong concordance between images within the CWRU/HIHG dataset and it appears that the algorithm can perform well in the presence of image quality variation. This is a necessity reflective of the repeated measure data that we present in the next chapter.

Upon examination of the quantitative results compared with the drusen area classification scheme developed by AREDS, we observed a steady increase in median drusen area when matched to their respective stepwise area classifiers. We do observe increasing variation as the drusen area category increases. This may be due in part to subjectivity in assignment of area grades by the AREDS retinal specialist since assessment is based on a mental calculation of area coverage using a set of referent circles of fixed size.

Chapter III

GENETIC ASSOCIATION ANALYSIS OF DRUSEN LOAD

Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in the aging population in the developed world. Current prevalence estimates for AMD in the United States are 1.52% and 16.39% for ages 70-74 and 80[23]. By 2020 more than 80 million people will have some form of AMD[96]. These prevalence estimates in the elderly population will continue to rise drastically as advancements in health care continue to increase average life expectancy.

AMD is a disease that results specifically in central vision loss as it affects the macula. Vision loss due to AMD is divided into “wet” and “dry” AMD[97]. “Wet” (neovascular) AMD is the result of new blood vessels arising behind the macula, which then begin to leak fluid into the space between the Bruch’s membrane and the RPE layer. Dry AMD is characterized by the presence of a broad range of abnormalities in the retinal pigment epithelial layer (RPE) and is considered more common in the overall number of AMD cases[98]. Abnormalities include the presence of drusen, hyperpigmentation or hypopigmentation of the RPE, and in the later stages of dry AMD progression, geographic atrophy. Although the role that drusen play in the pathogenesis of AMD is not currently known, manifestation of these yellow deposits between the RPE layer and the Bruch’s membrane is one of the hallmark clinical signs of AMD development[4]. Drusen can be characterized as hard or soft[91]. Hard drusen are typically less than 63 µm in diameter and have discrete borders. Soft drusen are commonly greater than 125 µm in diameter and typically have the property of fuzzy indistinct

borders. Many environmental factors have been attributed to AMD risk such as age, race, smoking, and cardiovascular risk factors such as obesity and hypertension[16, 20, 98-105]. In addition to the factors listed above, genetic studies have successfully identified common genetic variation in genes such as *CFH, HTRA1/ARMS2, C2/CFB* as well as 15 other loci in sample populations of European ancestry[50, 53-57, 61-63, 68, 106-110].

Of recent interest is the application of these environmental and genetic risk factors to create prognostic models of AMD risk[111-121]. Utilizing models that can predict AMD in the clinic may allow for an advantageous contribution to the field of precision medicine with respect to AMD treatment. One practical application of identifying the underlying risk factors for AMD is building models of disease progression. These models have focused on grouping samples broadly into progressors or non-progressors using a dichotomous endpoint, typically defined by measures of visual acuity. This approach does not take into account the importance of the rate of progression and uses a broad grading system rather than a quantitative measure of disease. By understanding how the environmental and genetic components of AMD contribute to the rate of progression, we have the potential to understand how to better tailor and administer treatment regimens and recommend more appropriate eye evaluation intervals.

Although significant contributions have been made in the treatment of wet AMD with the advent of anti-VEGF injections, there are currently few treatments available for dry AMD. Therefore, it is of substantial importance to develop models of AMD progression that emphasize characteristics common to dry AMD such as changes in drusen load over time. Here we examine the impact of a cumulative genetic risk score using 19 common AMD risk variants on drusen progression using data made available through the Age-related Eye Disease Study (AREDS), and a combined dataset from Case Western Reserve University (CWRU) and the University of Miami Hussman Institute for Human Genomics.

Materials and Methods

AREDS Dataset

AREDS was a major clinical trial that examined risk factors for AMD and cataracts, as well as prognostic factors and clinical course of disease. The impact of zinc and antioxidants on incidence and progression of AMD was also assessed. Phenotype data tables made available through the database of phenotypes and genotypes (dbGAP) on subjects who participated in the AREDS were examined for intermediate AMD without the presence of central geographic atrophy or neovascular AMD. Evaluation of the phenotype data was restricted to the 595 subjects that had longitudinal color fundus photographs deposited within dbGAP. Imaging data was available on subjects at 2 year intervals with a maximum of 12 years of follow-up. We selected subject eyes and that received consecutive diagnoses of intermediate AMD over a course of more than 2 years based on the AREDS dbGAP phenotype tables. This is represented by individual subject eyes receiving an AMD category of 3. An AMD category of 3 is represented by the presence of one of the following; one or more large drusen, greater than 20 averaged size drusen in the presence of soft drusen, 65 average-sized drusen in the absence of soft indistinct drusen, or non-central geographic atrophy. Imaging data was included up to the progression of intermediate AMD to a severe grade of central geographic atrophy or neovascular AMD, and selection of eyes for pulling out imaging data was carried out independently per subject eye. Extensive details about the AREDS grade categorization, study design, and subject information can be found in AREDS report 1[92]. Within dbGAP, 30-degree color fundus photos are available on 3 separate fields of the retina. Field 1M is centered on the temporal margin of the optic disc. Field 2M is centered on the macula, and Field 3M is centered temporal to the macula. For the

relevance of this study, JPG images were downloaded from dbGAP related to field 2M. Images were manually and computationally inspected for quality and poor images were removed from the dataset. The major features used in determining quality of the images Included in this the observation of images with poor focus, images that had the early treatment diabetic retinopathy study (ETDRS) grid still attached to the photo, no optic nerve-present within the photo, incorrect alignment of the camera leading to images containing differential color hues around the periphery of the fundus image, and also uneven illumination across the image.

HIHG Dataset and CWRU dataset

All cases were ascertained through the retinal clinics at the Vanderbilt Eye Institute (VEI) or the Bascom Palmer Eye Institute (BPEI) as part of a longitudinal study examining progression and response to treatment of severe AMD. 50-degree color fundus photos were taken at these visits using either a Zeiss 450 (Carl Zeiss, Jena, Germany) camera at the VEI, or a Topcon TRC 50IX camera (Topcon 50IA, Tokyo, Japan) at BPEI. Participants were graded by a retinal specialist on a 1-5 scale modified from AREDS at each visit with visit intervals ranging from 1, 4, 6,8,10, and 12 months[122, 123] . A grade of 1 or 2 was assigned to controls, grade 3 represented early/intermediate AMD, and grades 4 and 5 represented late AMD (geographic atrophy and choroidal neovascularization, respectively). Subjects were retrospectively examined for visits with color fundus photos graded as a 3 on the modified AREDS grading scale. Subjects that presented with continuous intermediate AMD for 1 year or more in the absence of geographic atrophy or neovascular-AMD had their imaging data examined. A minimum interval of 6 months between images was used in the CWRU/HIHG dataset[92, 122, 123]. In the situation that image intervals were less than 6 months apart, the higher quality image was selected from the range while still maintaining a 6 month separation between visits.

Drusen Quantification

Drusen quantification was completed using a previously developed automated drusen detection algorithm. Details about the detection algorithm are defined elsewhere but briefly described here[95]. Images are resized to a radius of 650 pixels for the non-black region of the image to obtain a standard resolution across all images. Anatomical structures are detected including the optic disc, fovea, vessels and image quality assessment for calculation of features for drusen detection. Drusen candidate pixels are extracted using a pixel classification algorithm. Drusen candidate regions are segmented using dynamic programming and drusen candidates are classified as being true druse using a large set of features including contrast and intensity changes within the image. Quantification is completed based on a threshold drusen probability map which is used to generate a binary drusen map to calculate drusen area. An ETDRS grid normalized as a 3000 μm radius from the fovea to the edge of the optic disc for each image is automatically placed. Although this is a fully automated process, each image was manually reviewed for proper placement of the ETDRS grid. We found that our digital ROI matched precisely with ETDRS grids left on digitized AREDS photos. Drusen surface area was quantified in each image's ROI as millimeters^2 and used for downstream analyses.

Genotyping

Genotyping data for both the CWRU/HIHG and AREDS datasets were made available through the International AMD Genomics Consortium (IAMDGC) (Fritsche et. al. 2015, In Review) and through permission of Dr. Emily Chew at the National Eye Institute (NEI). Details about the quality control procedures carried out by the IAMDGC can be found elsewhere (Fritsche et. al. 2015, In Review). Nineteen common variants previously associated with AMD were selected either through direct genotyping on the

IAMDGC exome chip array or through surrogate SNPs found to be in high LD (r2>0.8) with these variants if not directly genotyped.

Drusen Progression Rate Estimation

Since our data are represented by multiple visits per subject and drusen measurements in either one or two eyes, we employed a linear mixed effects model (LMEM) to estimate changes in drusen area over time using a similar modelling scheme to what has been presented for estimating GA progression rates[112]. Modeling of drusen progression rates was carried out using the R software package (R version 3.02) and the R-library "Linear mixed-effects models using Eigen and S4" (lme4, version 1.1-7). In brief, LMEM allows for an estimation of a population mean regression line known as a fixed-effect and deviations from that mean slope and intercept through estimation of random-effects. This multi-level modeling also allows for incorporation of information on both eyes for a subject if available, giving us a single measure of drusen growth. P-values were generated with the R-package "Tests in Linear Effects Models" (LmerTest,version 2.0-20) using Satterhwaite's approximations.

Analysis of Drusen Progression against Demographic, Endpoint Severity, and Treatment category

To examine the role of sex, smoking, AREDS treatment category, and age, the following statistical analysis plan was carried out. We first examined the impact of smoking, age, and treatment category independently within the mixed model of drusen progression to see the necessity of including these factors as covariates for the genetic analysis. Age was examined as a measure of correlation with drusen progression by way of a Pearson's correlation coefficient. Sex and smoking were examined as interaction terms within independent mixed-models. To understand the potential role of drusen growth on the progression of AMD from intermediate to neovascular AMD or geographic atrophy, study subject's endpoint severity was categorized into either

“drusen only”, “geographic atrophy”, or “neovascular AMD”. These severity categories were determined by using AREDS severity score information made available through the AREDS dbGAP data tables. Subjects were categorized based upon first instance of severe progression. In the case that subjects did not progress to geographic or neovascular AMD by the end of the full AREDS clinical trial and natural history study, subjects were categorized as having “drusen only”. Endpoint severity categories were coded as dummy variables and the “drusen only” category was set as referent within the model.

Cumulative Genetic Risk Score Analysis and Independent SNP Tests

For our primary analysis, the 19 variants or their respective surrogate SNPs were incorporated into a cumulative genetic risk score representing the major common variation contributing to AMD genetic risk[124]. Each variant was weighted by the effect size given by the AMD Gene consortium and multiplied by the number of risk alleles present at that locus. More details about the weighting scheme and application of this risk score can be found elsewhere[122, 124, 125]. To examine the effect of the genetic risk score on drusen progression, a Pearson’s correlation coefficient was calculated using each subject’s progression slope versus their cumulative genetic risk score. Examination of each of the common variants role in drusen progression was carried out independently by the inclusion of the number of risk alleles at each loci coded in an additive manner, and included as an interaction term within the mixed models.

Pathway Analysis

To examine the role of drusen progression in a pathway based analysis, a quantitative genome-wide association analysis was carried out in Plink using drusen progression rates as an outcome variable for the 274 AREDS subjects against the

genome-wide SNP data present on the exome-chip array[126]. A minor-allele frequency cutoff of 0.05 based on the 274 AREDS subject sample was used. Pathway analysis was performed on the results of the single variant tests using the Pathway Analysis by Randomization Incorporating Structure (PARIS) algorithm, and restricted to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, which contains 199 defined pathways[127].

Results

Demographics

Within the AREDS dataset, 246 of the 595 subjects available through dbGAP had either a baseline grade of intermediate drusen or progressed to intermediate drusen and had continuous visits of intermediate AMD for at least 2 years during the measured follow up (Table 3.1). In the CWRU/HIHG dataset, 75 subjects of the 500 subjects examined met the selection criteria as the independent dataset (Table 3.1). Analysis in the AREDS and CWRU/HIHHG datasets was restricted to the first 6.5 years of follow up and represents the median duration of the AREDS clinical trial. Although imaging data was available for the natural history portion of the AREDS study, a large proportion of the samples we selected for the study had progressed to severe AMD after the 6.5 year time point and thus did not have usable longitudinal data at that stage. The median number of visits with quantifiable images including baseline was 3.4 with a total of 973 images being used in the AREDS analysis and 75 subjects with a total of 272 images and a median number of visits of 3 for the CWRU/HIHG dataset. Of the 246 subjects AREDS subjects selected, 88 subjects had imaging data that fit our criteria in both eyes,

and 158 subjects had imaging data available for one eye. In the CWRU/HIHG dataset we observed 24 subjects with bilateral intermediate AMD. We observed high correlation in drusen area

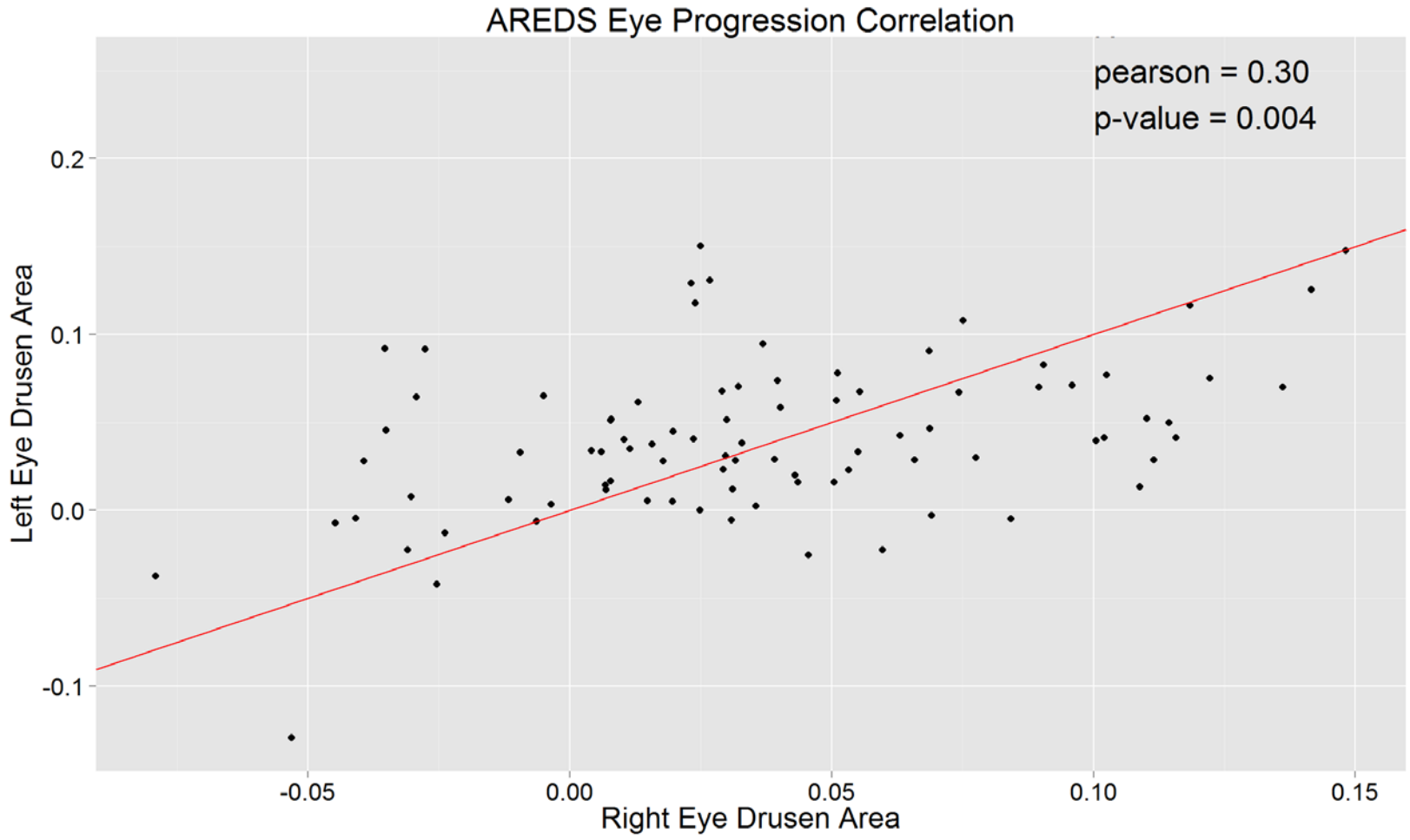
Table 3.1: Study Population Demographics
"AREDS Study" represents the entire sample set available in dbGAP with and without longitudinal imaging data. "AREDS Sampled" represents the 246 samples that had longitudinal data and met our filtering criteria. "CWRU / Miami" represents subjects with longitudinal imaging and genetic data available in this study population

Variable	AREDS Study (N = 4757)	AREDS Sampled (N = 246)	CWRU / HIHG (N=75)
Age at Exam (Years)			
Median	69.4	69.3	73.5
Gender, n (%)			
Male	2098 (44.1)	101 (41.1)	50 (66.7)
Female	2659 (55.9)	145 (58.9)	25 (33.3)
Ever Smoked, n (%)			
Yes	2650 (55.7)	151 (61.4)	26 (60.4)
No	2107 (44.3)	95 (38.6)	17 (39.6)

between eyes of subjects at baseline (rho = 0.857, p-value <2.20*10^-6; Figure 3.1) in bilateral intermediate AMD subjects. Correlation of bilateral progression was also significant (rho = 0.300, p-value = 0.004; Figure 3.2) within these 88 subjects. We observed significantly higher drusen area in subjects with both eyes present in the study versus subjects that had just one eye in our study (p-value = 0.024). We do not observe a significant association of drusen progression with sex, age at first exam, smoking status or treatment category.

Figure 3.1: Correlation at Baseline between Eyes

Figure 3.2: Correlation in Progression between Eyes



Genetic Data

We observe correlation with drusen baseline area and the cumulative genetic risk score (rho = 0.175, p-value = 0.006; Figure 3.3). We do not observe significant correlation between the 19 variant cumulative genetic risk score and drusen progression (rho = 0.039; p-value = 0.543; Figure 3.4). We do not observe a significant association with the top four highly associated variants reported by the IAMDGC and drusen progression (Figure 3.5 and Table 3.2). In single marker tests of the SNPs that make up the risk score, we observe a nominally significant association with rs943080 in VEGFA (p-value = 0.0281; Appendix A), this does not pass multiple testing correction. Pathway analysis was performed using the results of the quantitative GWAS based on the IAMDGC exome-chip array. In total, p-values were generated for 252,376 variants that met the minor allele-frequency cutoff of 0.05 and a genotyping efficiency of 95% (Appendix B). Of the 199 pathways that were interrogated as part of the KEGG database, the most highly associated pathway that passed Bonferonni correction was the cell adhesion molecules pathway (corrected p-value 0.02; Appendix C).

Figure 3.3: Correlation of Drusen Area at Baseline and Risk Score

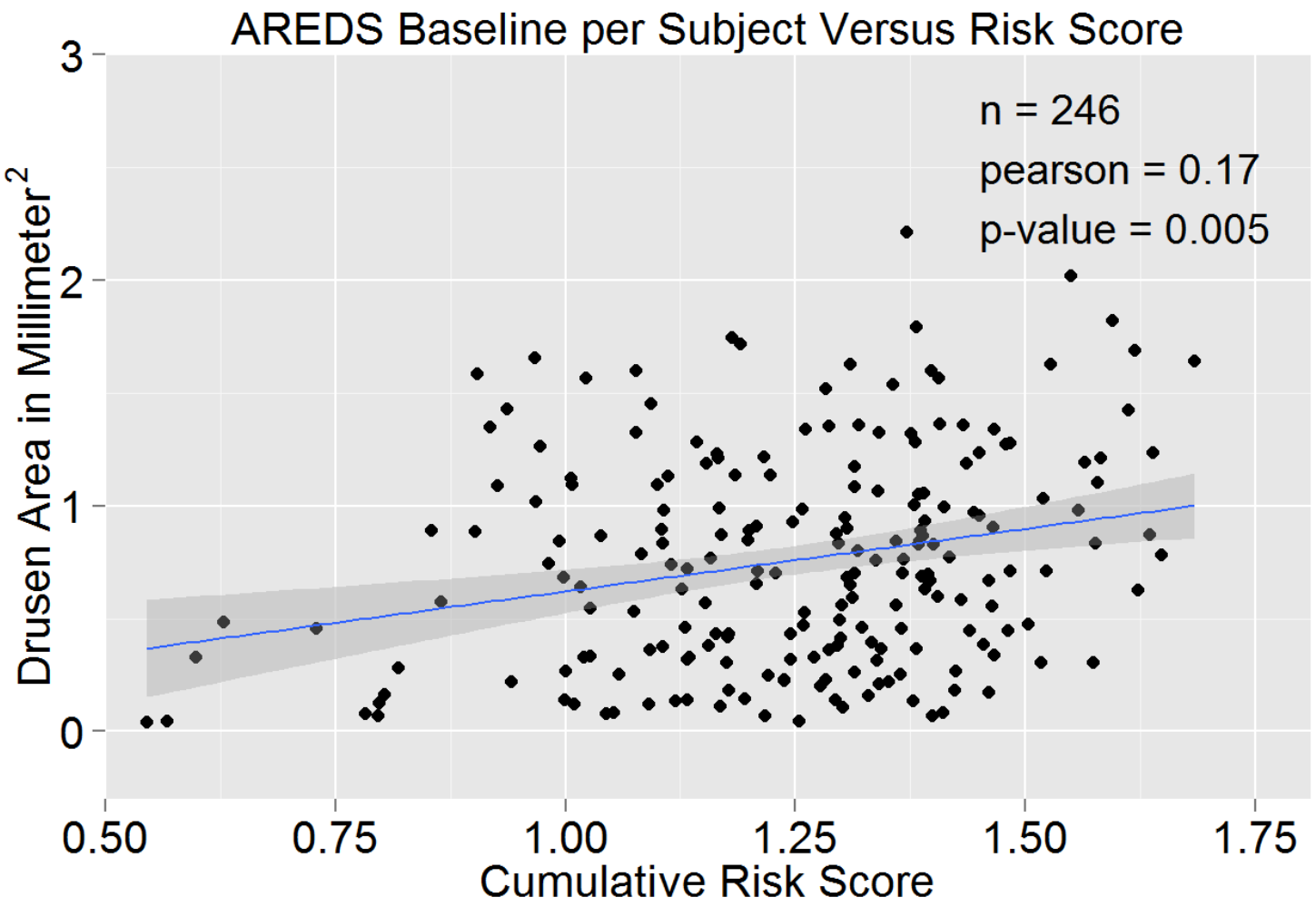


Figure 3.4: Correlation of Drusen Progression and Risk Score

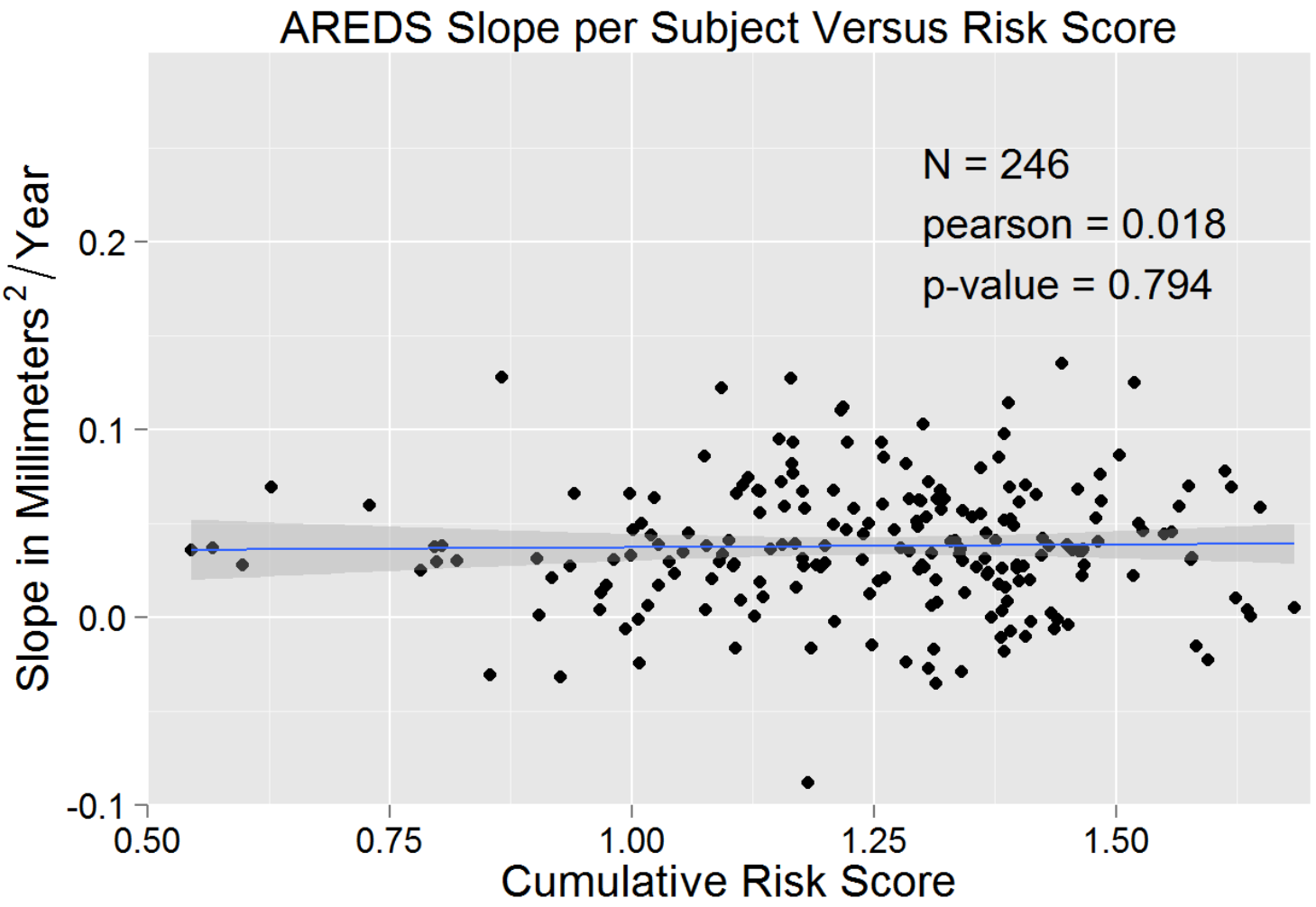


Table 3.2: Results of Single-Variant Association Analysis with Progression

	AREDS			CWRU / HIHG		
Variant	**Estimate**	**SE**	**P-value**	**Estimate**	**SE**	**P-value**
rs3812111_T_COL10A1	-0.01	0.007	0.193	-0.001	0.02	0.954
rs6795735_T_ADAMTS9.MIR548A2	0.003	0.008	0.681	-0.006	0.021	0.778
rs4698775_G_CFI	0.008	0.007	0.293	-0.019	0.018	0.308
rs9542236_C_B3GALTL	-0.012	0.007	0.099	0.023	0.023	0.314
rs920915_C_LIPC	-0.001	0.007	0.932	0.025	0.023	0.283
rs8017304_A_RAD51B	-0.015	0.008	0.06	0.024	0.02	0.246
rs943080_T_VEGFA	**0.016**	**0.007**	**0.028**	**0.01**	**0.018**	**0.581**
rs8135665_T_SLC16A8	-0.008	0.009	0.383	0.023	0.024	0.347
rs334353_T_TGFBR1	-0.013	0.009	0.126	0.031	0.023	0.198
rs79037040_T_tnfrsf10b_surrogate	0.009	0.007	0.216	0.021	0.024	0.375
rs115515129_G_ddr_surrogate	-0.0001	0.009	0.988	0.04	0.025	0.111
rs13081855_T_COL8A1.FILIP1L	-0.009	0.012	0.453	-0.015	0.043	0.736
rs1864163_G_CETP	0.006	0.009	0.46	-0.009	0.026	0.731
rs4420638_A_APOE	-0.006	0.01	0.579	-0.047	0.033	0.155
rs2230199_G_C3	0.003	0.008	0.728	-0.03	0.02	0.143
rs5754227_T_timp_surrogate	0.018	0.011	0.106	0.026	0.03	0.4
rs116503776_G_c2_cfb_surrogate	-0.005	0.014	0.712	-0.026	0.043	0.553
rs10737680_A_CFH	0.013	0.009	0.156	-0.015	0.026	0.556
rs10490924_T_ARMS2	-0.002	0.008	0.781	0.005	0.019	0.809

Figure 3.5 A-E: Drusen Area Progression by Genotype
Figure A. ARMS2 Risk Alleles in the AREDS dataset

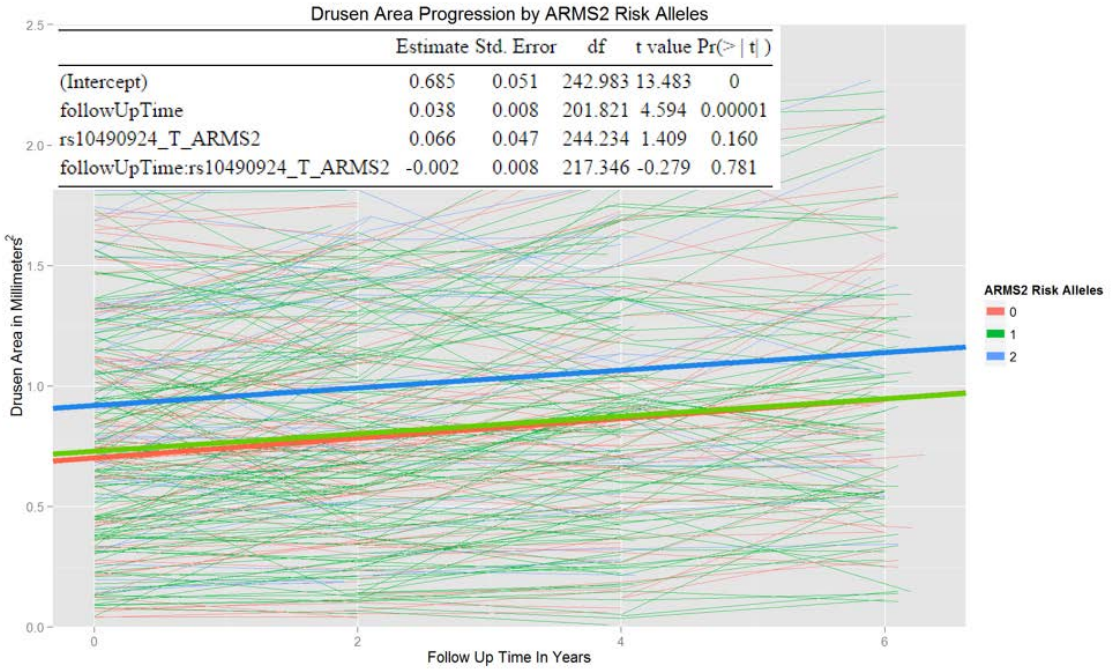


Figure 3.5 A-E: Drusen Area Progression by Genotype
Figure B. *CFH* Risk Alleles in the AREDS dataset

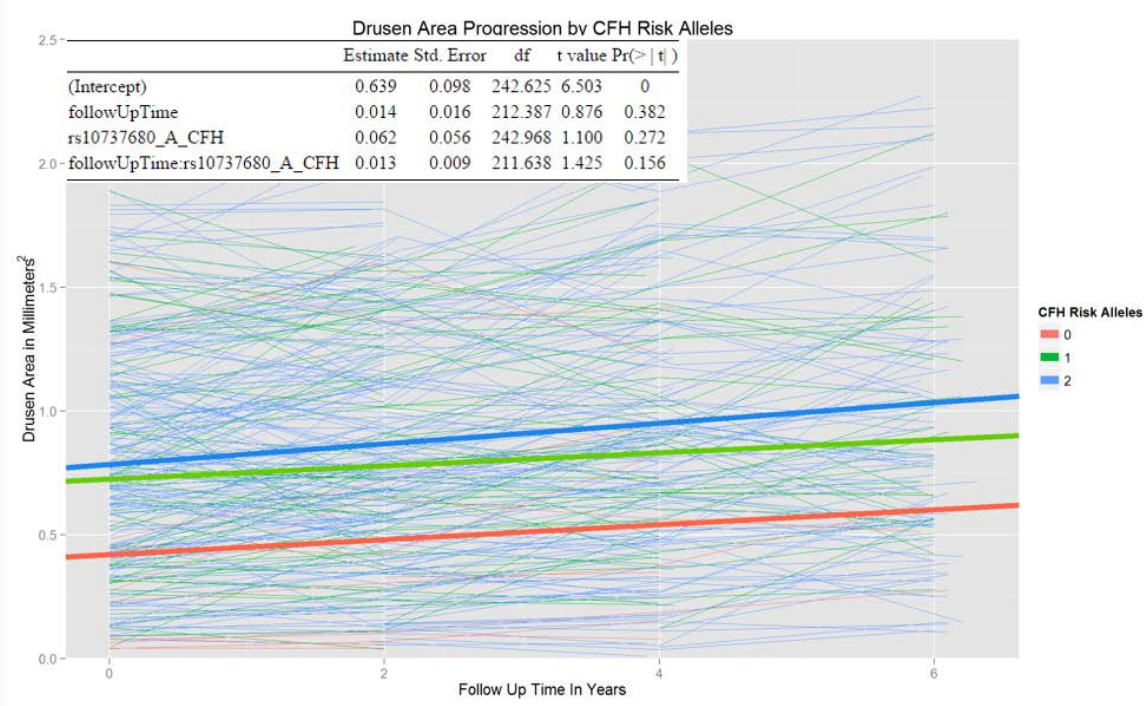


Figure 3.5 A-E: Drusen Area Progression by Genotype
Figure C. C2/CFB Risk Alleles in the AREDS dataset

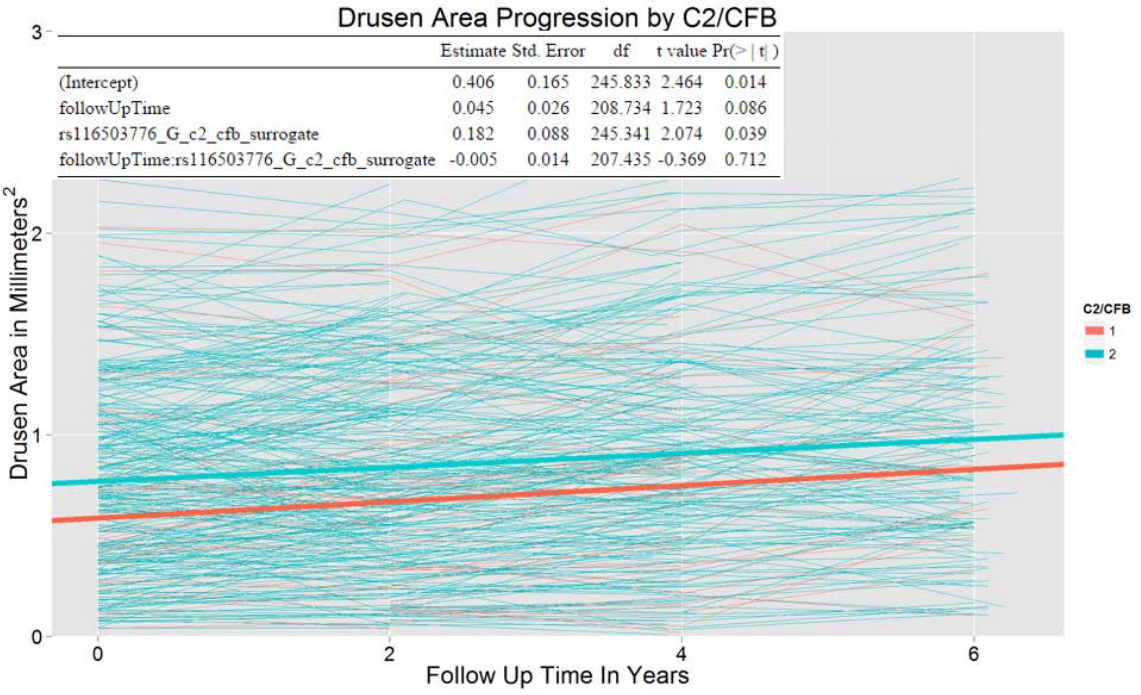


Figure 3.5 A-E: Drusen Area Progression by Genotype
Figure D. C3 Risk Alleles in the AREDS dataset

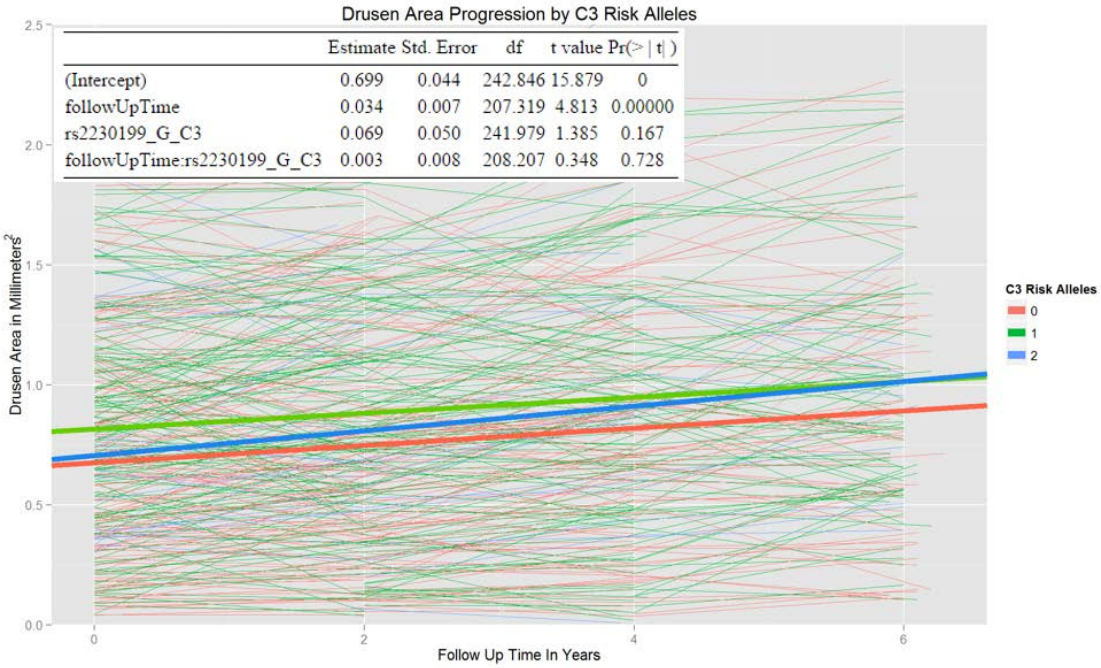
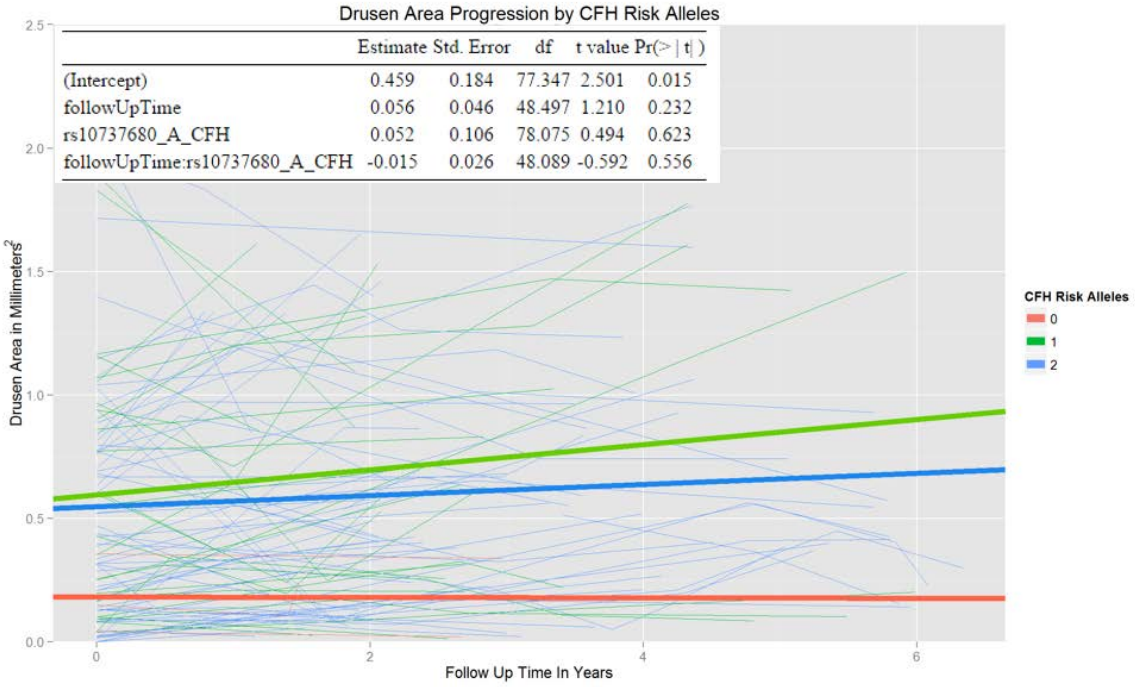


Figure 3.5 A-E: Drusen Area Progression by Genotype
Figure E. *CFH* Risk Alleles in the CWRU/HIHG Dataset



Discussion

AMD is a complex disease containing phenotypic heterogeneity with respect to the combinatorial presence of drusen, geographic atrophy, and neovascular AMD. In this analysis we examined the potential role of the 19 major genetic risk loci to drusen growth during the intermediate stages of the disease in the absence of neovascular AMD and geographic atrophy. Since the bulk of studies examining drusen have treated this phenotype as a binary response, we wanted to potentially refine the precision of examining this phenotype by treating it as a quantitative variable through a previously established drusen quantification algorithm that takes advantage of color fundus photos[95].

In a recent study performed by the AREDS study research group it was found that increasing drusen severity at baseline was a significant predictor for progressing to geographic atrophy and neovascular AMD[92]. We observed similar findings when examining drusen at baseline as a continuous variable and the 10 year endpoint severity outcome of study participants that met our inclusion criteria. Study eyes that maintained intermediate AMD throughout the course of the AREDS study had significantly lower average total drusen area compared to subjects that had baseline intermediate and progressed to geographic atrophy and neovascular AMD. This observation is present in the CWRU/HIHG dataset as well. We did not observe significant differences in rate of drusen progression within the first 6.5 years of follow up and the 10 year outcome of severity.

Although smoking and age are major risk factors for AMD they do not appear to have a major effect on drusen within our dataset. It is important to note that we do not differentiate subjects based on whether they were current smokers during the time of the

AREDS study or by pack years of smoking exposure. Age was limited to those aged 55 years or older at baseline and subjects that already have presence of intermediate AMD. We observe a highly significant correlation of drusen area within subjects that present with bilateral intermediate AMD (rho = 0.847, p-value < 0.0001). This observation may be inflated as we are not including subject eyes that have either severe AMD in the fellow eye, or not enough medium to large drusen to be classified as intermediate AMD. These correlation findings have also been observed by other groups looking at bilateral drusen using optical coherence tomography and thus our results are consistent with these previous findings[128]. When examining correlation in bilateral drusen progression within these same subjects, we see significant although reduced correlation (rho = 0.300, p-value = 0.004).

Genotype data were chosen based on previous work carried out by the AMDGENE consortium that identified 19 major common AMD risk variants that may explain up to 65% of the variation seen in AMD[124]. For our primary genetic analysis, these 19 variants were aggregated into a cumulative genetic score to examine its impact on drusen at baseline and progression. The risk score analysis revealed significant correlation with drusen area at baseline (rho = 0.17, p-value = 0.005) and no correlation with drusen progression (rho = 0.039; p-value = 0.543). It is important to note that SNP weights assigned to the variants used in the risk score are based on the AMDGENE consortium analysis that performed a cross-sectional analysis of risk and as such these weights may not be representative of their role in progression. In a secondary analysis we attempted to dissect the role of these variants to drusen progression but did not find any single variant that substantially contributes to drusen progression.

As we are presenting a case-only analysis in 2 smaller datasets, power is a significant concern. Post-hoc power calculations show that we had had 80% power to detect an effect size of 0.0196mm^2/year when correcting for the 4 major loci. We had

56% power to detect a significant association in the cumulative genetic risk score progression analysis based on the estimated effect size and sample size in the AREDS dataset. Power in the independent association analysis for the big four variants within *CFH*, ARMS2/HTRA1, C2/CFB, and C3 was also severely limited at under 10% after multiple test correction. Variant rs943080, which is near *VEGFA*, was the only variant nominally associated with drusen progression out of 19 loci that make up the cumulative genetic risk score. Power to detect an association in this variant was 37% based on 19 test corrections. This variant did not replicate within our Miami / CRWU dataset although the sample size and power of this dataset was even further limited.

A second exploratory approach is to aggregate potential effects not using a genetic risk score, but on functional relatedness. We performed a pathway based analysis to see whether any functional pathways within the KEGG database were enriched for drusen progression. The top pathway that was enriched was the cell adhesion molecule pathway (KEGG database id: hsa04514; Appendix C). We identified three genes driving the signal. These genes were *neurofascin* (*NFASC*) on chromosome 1 (p-value = 0.0004), *CD226* molecule (CD226) on chromosome 18 ($p < 0.0002$), and *neurexin 1* (*NRXN1*) on chromosome 2 (p-value = 0.0006). The statistical significance of these genes was confirmed using the gene enrichment program "Versatile Gene-based Association Study" (VEGAS). Results from the VEGAS analysis can be found in appendix D. Molecules in this pathway play a role in a wide array of functions including inflammation and immune response. Previous work has shown that inflammation between the RPE layer and the Bruch's membrane may play a role in AMD associated drusen formation mainly through cellular debris trapped between these layers[3, 129]. In another study examining the impact of cell adhesion molecules on AMD, it was shown that soluble vascular cell adhesion molecule 1 associates with increased incidence of early AMD[130]. These findings highlight that although the major 19 risk loci may not

contribute significantly to drusen progression within our study, there may be other variants of functional importance directly or indirectly impacting drusen growth during the intermediate stage of disease.

One of the major limitations that we found during the course of this study was the quality of images available from the AREDS dbGAP dataset. AREDS images made available in dbGAP are not original images, but digitized copies of slide transparency film. Photographs taken at the VEI and BPEI were native digital images. All images from both the native digital images obtained in the CWRU/HIHG dataset and the digitized photos from AREDS were assessed for quality using an automated algorithm that assigns a quantitative quality score to the image based on a number of metrics including blurriness, contrast in the image, and ability to delineate blood vessels surrounding the macular. A major proportion of the AREDS images were considered poor by this algorithm, which did not necessarily reflect the qualitative assessment of the images by a retinal specialist. Assessing the quality of the directly digitized images from the Miami/Case Western data we observed most images falling into the high quality range both quantitatively and qualitatively. It may be that the process of taking the fundus slides from transparent film in the AREDS dataset to digital copies affects the ability for accurate automatic quality assessment of the images. We must also be aware that the overall low quality of the AREDS images when examined in a longitudinal format may be introducing too much variation across time-points to accurately measure drusen progression. This may be somewhat reflected between differences in correlation between intra-subject drusen at baseline and bilateral drusen progression, where we see a highly correlated baseline measurement, but reduced bilateral progression correlation. Thus if the effect sizes of the major common risk loci are more modest for drusen growth , we may not be able to overcome the signal to noise ratio introduced by image quality. Although the CWRU/HIHG images were of superior quality, the size of the dataset

limited our power to detect a true association with drusen at baseline or drusen progression. Although we were able to retrospectively collect previous retinal visits on subjects within the progression and treatment study, the major ascertainment criteria for these participants was severe AMD, thus longitudinal data was sparse prior to the presence of reduction in visual acuity.

It is important to note that no study to date has been published that details the role of the major AMD risk loci in quantitative drusen progression, and thus is an important step in understanding the role of drusen etiology in AMD genetics.

Chapter IV

RARE COMPLEMENT FACTOR H VARIANT ASSOCIATED WITH AGE-RELATED MACULAR DEGENERATION IN THE AMISH[1]

Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in individuals over the age of 65 in the developed world[23, 131]. AMD is a progressive neurodegenerative disease that results in central vision loss. Vision loss caused by AMD is generally divided into two categories. Non-neovascular or "dry" AMD is characterized by the presence of a broad range of abnormalities in the retinal pigment epithelial layer (RPE). Clinical characteristics include the presence of drusen, hyperpigmentation or hypopigmentation of the RPE, and, in later stages, geographic atrophy, which results from retinal pigment epithelial (RPE) layer cell death[132]. Neovascular or "wet" AMD occurs when new blood vessels form behind the macula, leaking fluid into the space separating the Bruch's membrane and the RPE layer[132].

AMD risk has been attributed to many lifestyle influences such as age, race, smoking, and cardiovascular risk factors including obesity and hypertension[20, 98, 103, 107, 133]. Genetic studies have identified common variants with strong associations in *CFH* and *ARMS2/HTRA1*, as well as multiple loci of smaller effect, in populations of European

[1] Adapted from: Joshua D. Hoffman, Jessica N. Cooke Bailey, Laura D'Aoust, William Cade, Juan Ayala-Haedo, Denise Fuzzell, Renee Laux, Larry D. Adams, Lori Reinhart-Mercer, Laura Caywood, Patrice Whitehead-Gay, Anita Agarwal, Gaofeng Wang, William K. Scott, Margaret A. Pericak-Vance, Jonathan L. Haines. *Investigative Ophthalmology & Visual Science*. 2014 June 6; 55(7):4455-60.

descent[46, 50, 55, 57, 62, 134]. A recent meta-analysis completed by the AMDGene Consortium showed that the most strongly associated variants in *CFH* and *ARMS2/HTRA1*, in combination with 17 other loci that reached genome-wide significance, account for less than 65% of the total genetic contribution to AMD[135]. While this shows success in common variant analysis, there is still a gap in the total genetic variation yet to be explained by this complex disease. Some of this missing heritability is thought to lie within rare variants of large effect[81, 136-138]. To address the deficit in knowledge of rare variants influencing AMD, we exome sequenced individual members of a nuclear family who represent a subset of the Ohio-Indiana Amish population.

The Amish are a genetically and culturally isolated founder population descended from Swiss and German Anabaptists who emigrated from Western Europe to North America in the 1700s and 1800s[139]. Individuals of the Amish community typically marry within the faith and observe a strict lifestyle, resulting in a community that is more genetically and environmentally homogeneous than the surrounding population. Due to the intermarriage within the community and relatively recent founder event, this population may also be enriched for some rare variations. These factors, in addition to an extensive family record available through the Anabaptist Genealogy Database (AGDB), make the Amish a unique and valued population for genetic studies[85, 140, 141].

Materials and Methods

Subjects:

Amish subjects were selected from the Collaborative Aging and Memory Project (CAMP), an ongoing sample collection of Amish individuals living in the United States in Ohio and Indiana, described elsewhere[87, 142]. Construction and maintenance of the Anabaptist Genealogy Database (AGDB) is covered under an IRB-approved protocol at the National Institutes of Health (Dr. Leslie Biesecker, Principal Investigator). AMD affection status was assigned based on self-report questionnaire response where subjects were asked if they had ever been diagnosed with AMD by a physician. A subset of 73 participants (42 cases, 31 controls) received a follow up clinical exam by a retinal specialist. The self-report dataset included 128 individuals with AMD, 728 individuals without AMD, and 294 with no self-report information, and all individuals were connected into a single 13-generation pedigree based on an "all common paths" query of the AGDB using PedHunter 2.0 software and shown in figure 4.1[143].

Non-Amish subjects were ascertained from the Duke University Eye Center (DUEC), the Vanderbilt Eye Institute (VEI), and the Bascom Palmer Eye Institute (BPEI) at the University of Miami Miller School Of Medicine. Participants were examined by a retinal specialist and graded on a severity scale derived from the Age-Related Eye Disease Study (AREDS), described elsewhere[40, 122]. Grades were given on a scale of 1-5 where grades 1 and 2 were assigned to controls, grade 3 represented early AMD, and grades 4 and 5 represented late AMD (geographic atrophy and choroidal neovascularization, respectively). Amish participants who received a clinical diagnosis were graded using the same criteria. The final non-Amish dataset contained 1,732

cases, 943 controls, and 310 unknown samples. Demographic information for both the Amish and non-Amish datasets is depicted in table 4.1.

Table 4.1:Demographics for Amish and Non-Amish Datasets

	Median Age at Exam [IQR]	Percent Female
Amish		
Cases	82 [80,86]	70%
Controls	80 [76,84]	57%
Non-Amish		
Cases	78 [72,83]	63%
Controls	70 [64,76]	58%

IQR: Interquartile range

As part of ongoing AMD family studies, a single nuclear Amish family (family 1) with multiple individuals affected with AMD (three of eight siblings affected) was noted. Affected members lacked risk alleles at the Y402H locus in *CFH* and at the A69S locus in *ARMS2*. The three affected siblings (a-78, a-79, a-80) (Figure 4.2) were clinically evaluated to confirm disease status. The other siblings were not available for clinical examination at the time of exome sequencing, although a self-report diagnosis was completed. Subject a-78 was diagnosed with bilateral choroidal neovascularization (Grade 5); subjects a-79 and a-80 were diagnosed with large drusen in both eyes (Grade 3). Details of clinical grades and age at time of diagnosis for the ascertained nuclear family are included in figure 4.2. All procedures followed the tenets of the Declaration of Helsinki and were approved by the institutional review boards of the University of Miami Miller School of Medicine and Vanderbilt University. Informed consent was obtained from all research subjects involved in this study.

Cumulative Genetic Risk Score Analysis:

Genetic risk scores were calculated using the 19 variants and their effect sizes reported by the AMDGene consortium's meta-analysis[15]. Genotyping was performed using the Sequenom MassARRAY genotyping platform (Sequenom, San Diego, USA). Subjects that presented with missing genotypes at any of the 19 loci were excluded from the analysis. Each variant was weighted and multiplied by the number of risk alleles present at each locus, where the SNP weight (*w*) is equal to the individual SNP beta-estimate divided by the sum of the beta-estimates across all 19 loci (formula 1). Risk scores from each of the 19 loci were summed to give the cumulative genetic risk score per person.

Figure 4.1: Diagram of the entire 13 generation all connecting pedigree of the Amish of Indiana and Ohio.

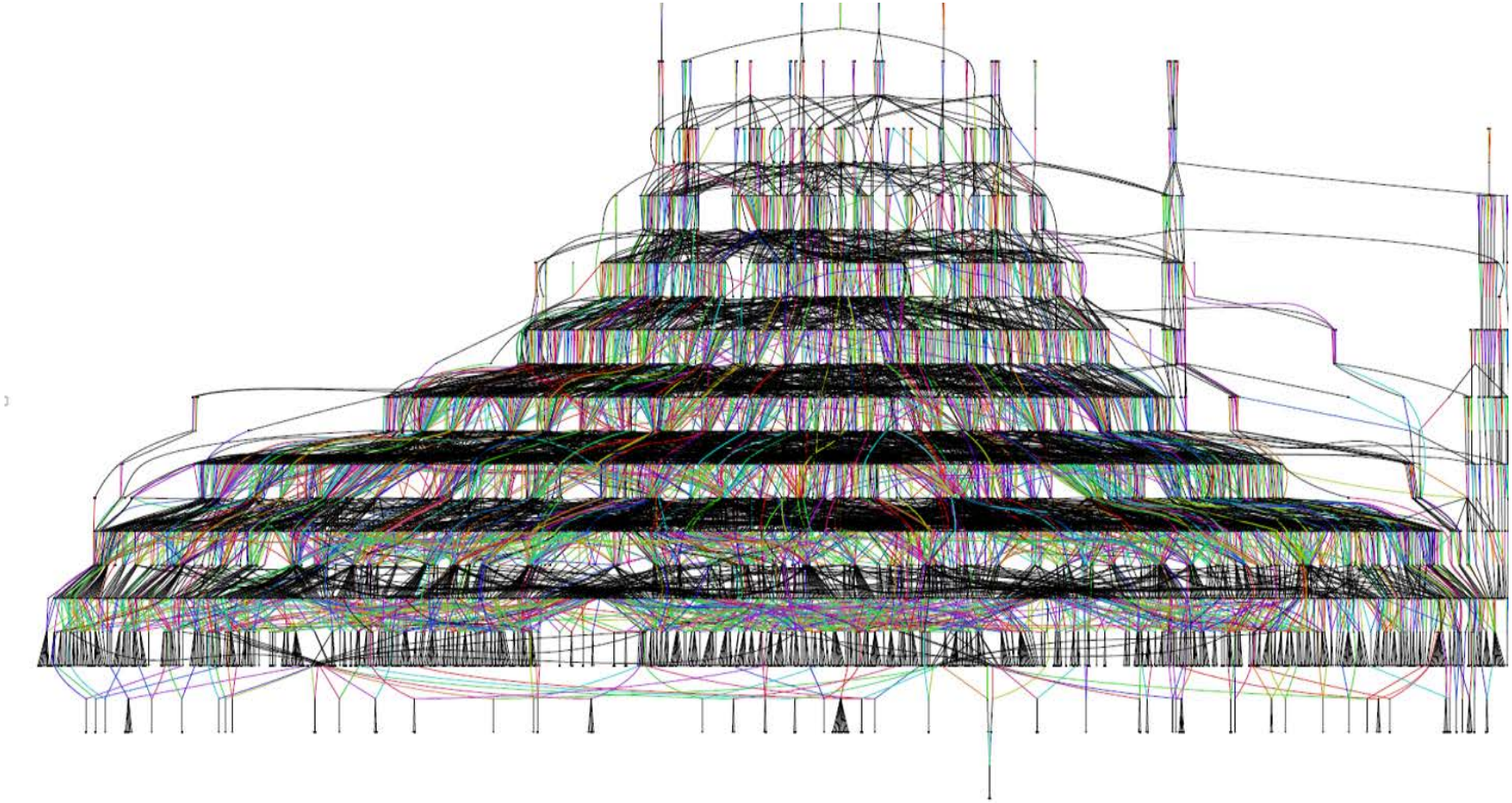
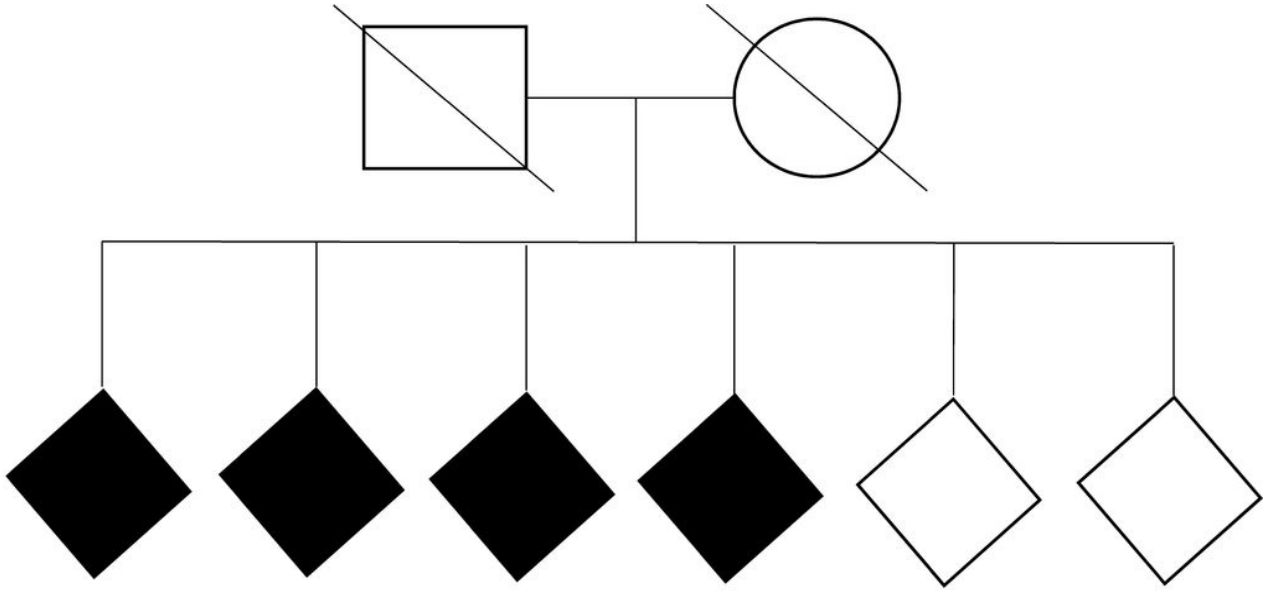


Figure 4.2: Pedigree of nuclear family chosen for exome sequencing.

*Clinical diagnoses were available for a-78, a-79, and a-80 at time of exome sequencing. Grading was carried out according to the modified AREDS scale.



Subject	a-78	a-79	a-80	a-81	a-82	a-83
Exome Sequenced	+	+	+			
OD/OS DX*	5b/5b	3a/3a	3a/3a	3a/3a	1a/2a	1b/1b
P503A Carrier	+	+	+			+
A69S Carrier						+
Y402H Carrier				+		
Age at Exam	75	72	68	74	71	72

Two-sided t-tests assuming unequal variance were calculated in the Amish case-control group and Non-Amish versus Amish case group.

(1) Cumulative genetic risk score = w_1 *Genotype*$_1$ + w_2 *Genotype*$_2$ + … + w_{19} *Genotype*$_{19}$

Exome Sequencing:

DNA for all samples was extracted from whole blood by the Vanderbilt University DNA Resources core and the John P. Hussman Institute of Human Genomics (HIHG) at the University Of Miami Miller School Of Medicine using PureGene DNA extraction methods (Gentra Systems, Minneapolis, MN). The three clinically examined members of family 1 were selected for exome sequencing. Exome capture was performed using the Agilent SureSelect All Exon kit V. Exon enriched libraries were sequenced on the Illumina HiSeq 2000. Sequence capture and high-throughput sequencing were completed at the HIHG. Paired end reads were generated and mapped to the human reference genome (version hg19 from UCSC) using Burrows-Wheeler Aligner[144]. Duplicate reads were marked using Picard tools and local realignment around insertions and deletions was performed using the Genome Analysis Toolkit (GATK)[145]. Realignment was performed using the base quality score recalibration walker followed by variant calling. Variant quality score recalibration was completed using an additional 172 sets of exomes from Amish individuals who were not members of nuclear family 1. Annotation of nucleotide variants was performed using the SeattleSeq Annotation server[133]. Variant filtration was completed as follows: variants found in the 1000 Genomes project database, NHLBI exome sequencing project exome variation server (EVS) database, or dbSNP137 database were excluded. In addition, single nucleotide variants that were not missense, nonsense, splice junction, or frame shift causing mutations found within exon boundaries were excluded. Genes not known to be associated with AMD were also excluded. Variants found to be either homozygous or heterozygous and shared by all three affected siblings were retained**.** Variant

confirmation was performed using standard forward and reverse Sanger sequencing practices on the ABI 3730xl.

Targeted Genotyping:

Genotyping was performed using the Sequenom MassARRAY genotyping platform (Sequenom, San Diego, USA). Independent Sequenom pools were genotyped and evaluated for the rare-variant data, and the 19 loci used to calculate the cumulative genetic risk score. A genotyping efficiency threshold of 95% was used in both datasets to determine a valid Sequenom assay. See table 4.2 for final sample sizes used in each step of the analysis.

Association and Linkage Analysis:

Variants identified in the nuclear family were evaluated for association with AMD in the full Amish dataset using the Modified Quasi Likelihood Score statistic (MQLS)[146]. MQLS is comparable to the chi-square test with the exception that it estimates the variance on point estimates of the allele frequencies in cases and controls while taking into account the correlation between related individuals in a pedigree. This effectively allows all degrees of relationships to be included in the association analysis. Kinship and inbreeding coefficients for all possible relationships in the 13-generation pedigree were calculated using the MQLS recommended program KinInbcoef. MQLS analysis was performed using option 1, which allows for individuals with genotype, but no phenotype data to contribute to the analysis. An assumed disease prevalence of 10% was selected from published estimates of the prevalence of AMD in individuals over age 60 in population samples[23, 131, 147]. The nuclear family was genotyped as part of the full Amish dataset and was not excluded from the analysis. A sub-analysis was performed in which only those cases with a clinical exam completed by a retinal specialist were analyzed.

Parametric linkage analysis was performed using the entire data set genotyped on the Affymetrix Human SNP Array 6.0. The pedigree was divided into smaller, more

Table 4.2: Genotyped Samples Utilized per Analysis Step

	Full Dataset	Linkage-Scan	Exome Sequencing	Rare Variant Association	Risk Score Analysis
Amish					
Cases	128	95	3	95	88
Controls	728	1	-	653	559
Unknowns	294	-	-	225	-
Non-Amish					
Cases	1732	-	-	1456	1573
Controls	943	-	-	791	841
Unknowns	310	-	-	124	-

Sample numbers shown are post quality control

computationally feasible pedigrees using PedCut with a bit size threshold of 24[148]. Parametric heterogeneity log-of-odds (HLOD) scores were calculated under affecteds-only dominant and recessive models assuming incomplete penetrance in MERLIN[149]. Under the dominant model, we specified penetrance values of 0 for no copies of the disease allele and 0.0001 for one or two copies of the disease allele. Under the recessive model, we specified penetrances of 0 for zero or one copy of the disease allele and 0.0001 for 2 copies of the disease allele. The disease allele frequency was set to 10% and marker allele frequencies corrected for relatedness were estimated from all genotyped Amish individuals. Multipoint linkage analysis was performed on a 7 megabase region surrounding *CFH*. LD pruning for multipoint was performed using the HAPMAP CEPH samples with a pairwise r^2 cutoff of < 0.16. In addition to standard multipoint linkage analysis, sub-analyses specifying liability classes modeled on the effect sizes of the rare and common *CFH* alleles were examined. These included using the common SNP data derived from the Affymetrix 6.0 GWAS chip; assuming only *CFH* Y402H as a risk allele; assuming only *CFH* P503A as a risk allele, and assuming both *CFH* Y402H and P503A as risk alleles. Detailed methods for quality control and linkage analysis procedures can be found elsewhere[150].

To examine linkage disequilibrium (LD) between the novel variant and the common Y402H variant, we extracted the most distantly related Amish subjects genotyped in the pedigree. For this analysis, we used a maximum pair-wise relationship cut-off between 2nd and 3rd cousins, resulting 168 individuals available for an LD calculation. LD was measured using the software package PLINK[126].

Results

Clinical vs. Self-Report of AMD:

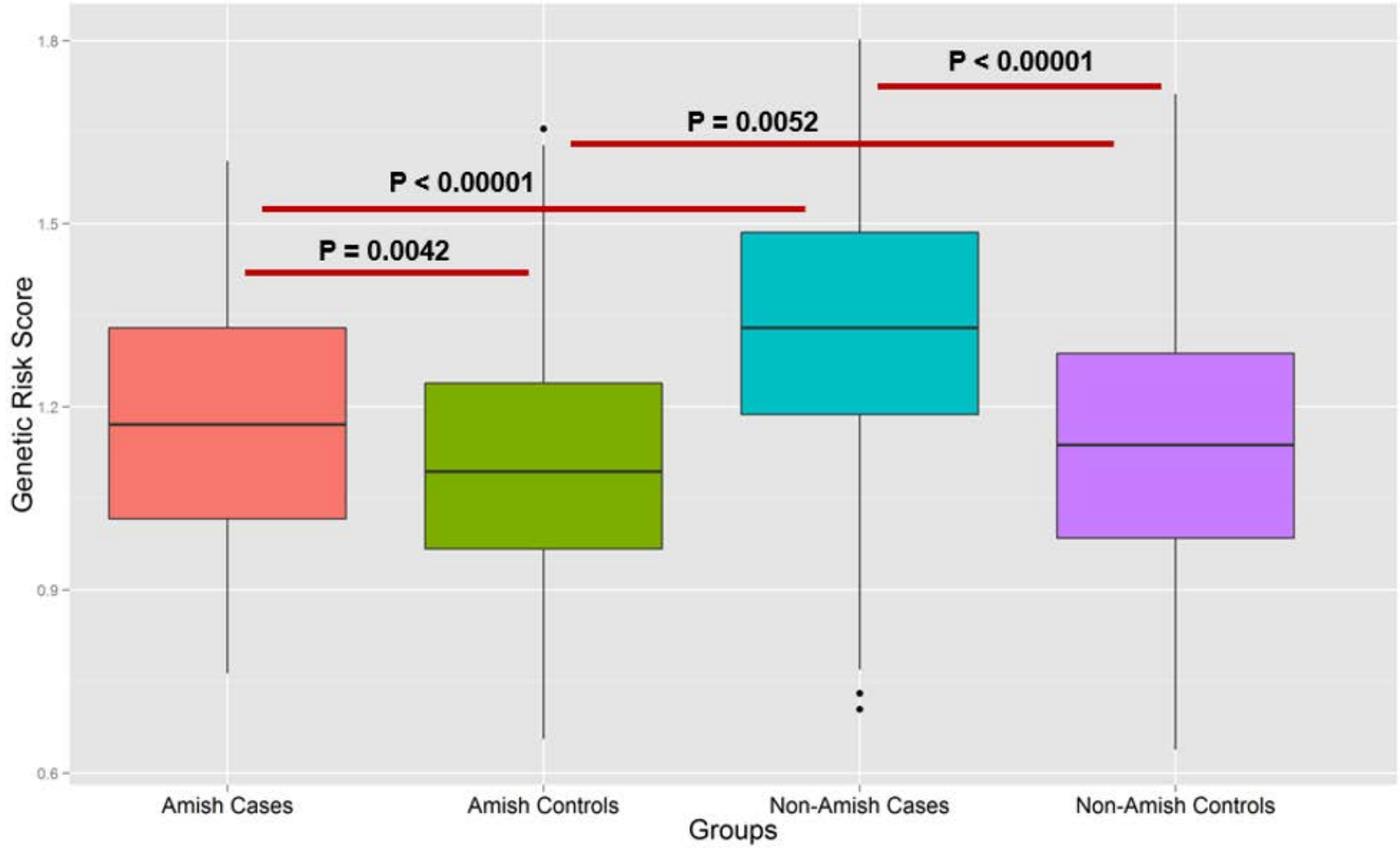
Comparing clinical diagnoses to self-report of AMD status in individuals for whom both were available, we observed positive and negative predictive values of 89% and 90%, respectively (Table 4.3), indicating that self-report of AMD status is a good proxy for AMD diagnosis in this population sample.

Cumulative Risk Score Analysis: In our cumulative genetic risk score analysis we observe a mean risk score of 1.12 (95% CI [1.10, 1.13]) in the Amish controls and 1.18 (95% CI [1.13, 1.22]) in the Amish cases (p=0.0042; Figure 4.3). We also observe a mean risk score

Table 4.3: AMD Self-report Pilot Study

Clinical Exam	Self-Report AMD	Self-Report No AMD	Total
Diagnosed with AMD	39	3	42
No AMD	5	26	31
Total	44	29	73

Figure 4.3: Cumulative Genetic Risk Score Analysis Across Amish and non-Amish Samples



of 1.14 (95% CI [1.13, 1.16]) in 841 non-Amish Caucasian controls and 1.31(95% CI [1.30, 1.32] in 1,573 non-Amish Caucasian cases. When comparing the Amish cases with the non-Amish cases we see a significant decrease in genetic risk score ($p<0.00001$; Figure 4.3).

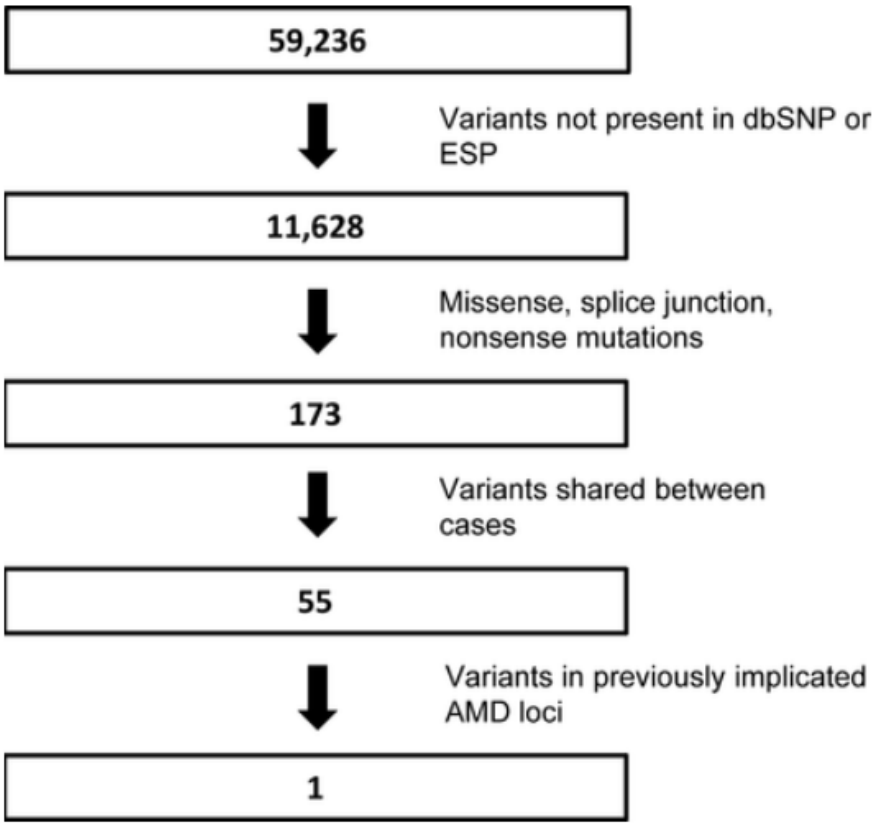
Exome Sequencing and Linkage Analysis:

Exome sequencing was performed on the three individuals in family 1. We generated on average 4.6 million reads per sample. 75% of reads were on target with a depth of coverage of 10X or higher. We examined these data for rare variants in known AMD genes that might explain AMD in this family. After variant filtration procedures, we identified a single non-synonymous mutation in *CFH* that predicts a proline to alanine amino acid change at position 503 (P503A; Figure 4.4).

Case-control analysis using self-reported affection status in the Amish sample population identified a significant association of AMD with P503A ($p=9.27 \times 10^{-13}$). Results were consistent in the sub analysis of subjects with clinically confirmed AMD ($p=5.21 \times 10^{-7}$). Out of the Amish samples that were not part of the original nuclear family and that were genotyped for the P503A variant, we observed 15 additional carriers of the P503A variant with 8 subjects affected, 5 unaffected, and 2 subjects of unknown case-control status. When evaluating the variant in our non-Amish Caucasian dataset, we did not observe the risk allele in either the 791 controls or the 1,456 cases. Multipoint linkage analysis carried out on chromosome 1 identified an HLOD peak of 5.12 on chromosome 1 spanning the *CFH* gene. After Incorporating a liability class for carriers of the *CFH* P503A rare variant we observe a maximum LOD score of 4.53 in this region; by including the common Y402H variant we observe a maximum LOD score of 3.72, when we include liability classes for carriers of both the common and rare *CFH* variants into

the model, we observe a maximum HLOD score of 3.28 (Figure 4.5 A-D). Upon examination of LD between the Y402H variant and the P503A variant we observe an r^2 value of 0.002.

Figure 4.4: Variant filtration procedures
Fiftration steps used and the resultant number of variants after each stage. Variants passing quality control procedures are used as the initial filtration starting point. ESP refers to the NHLBI grand opportunity exome sequencing project.



Discussion

We determined that the genetic burden of known AMD loci is substantially lower in the Amish than in the general European ancestry populations. Given that AMD is at least as frequent in the Amish as in other European ancestry populations (Dwight Stambolian, personal communication) and that the Amish generally do not smoke (the strongest known AMD environmental risk factor outside of age), our data support the hypothesis that other genetic loci are segregating in the Amish.

We identified a densely affected nuclear Amish family in which affected siblings a-78, a79 and a-80 do not carry the Y402H or A69S risk variants in *CFH* and *ARMS2* respectively, loci that account for the majority of the genetic risk of AMD in Caucasian populations[107]. The absence of these risk alleles in affected members lead us to hypothesize that other rare variants of large effect may be contributing to AMD in this family. Using exome sequencing data, we identified a novel missense mutation that is shared among the affected siblings and located in the *CFH* gene. This mutation is a cytosine to guanine transversion resulting in the substitution of an alanine for a proline at amino acid position 503 (P503A). The P503A locus has a Genomic Evolutionary Rate Profiling (GERP) score of 3.64, indicating strong conservation across mammalian species. PolyPhen2, which predicts the possible impact of amino acid substitutions on the structure and function of human proteins, indicates that this variant is probably damaging.

Figure 4.5 A-D: Multipoint linkage Results on Chromsome 1

Figure A. Native multipoint linkage results in region harboring *CFH*

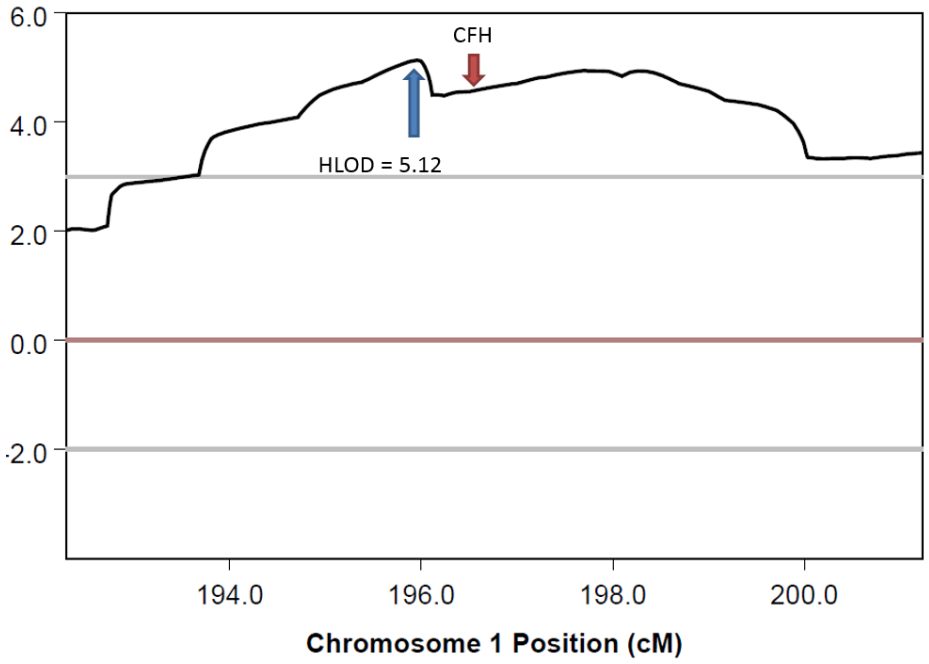


Figure 4.5 A-D: Multipoint linkage Results on Chromsome 1

Figure B. Incorporation of the P503A variant as a covariate in the linkage analysis shows a reduction in the Maximum LOD score from 5.12 to 4.53.

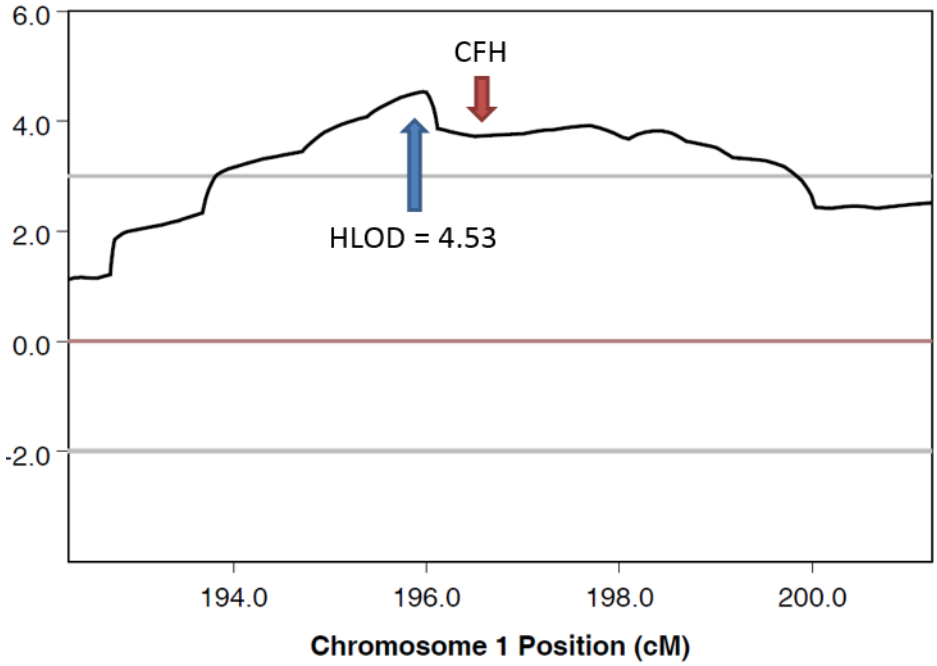


Figure 4.5 A-D: Multipoint linkage Results on Chromsome 1

Figure C. Incorporation of the P503A variant as a covariate shows a reduction in the Maximum LOD score from 5.12 to 3.72.

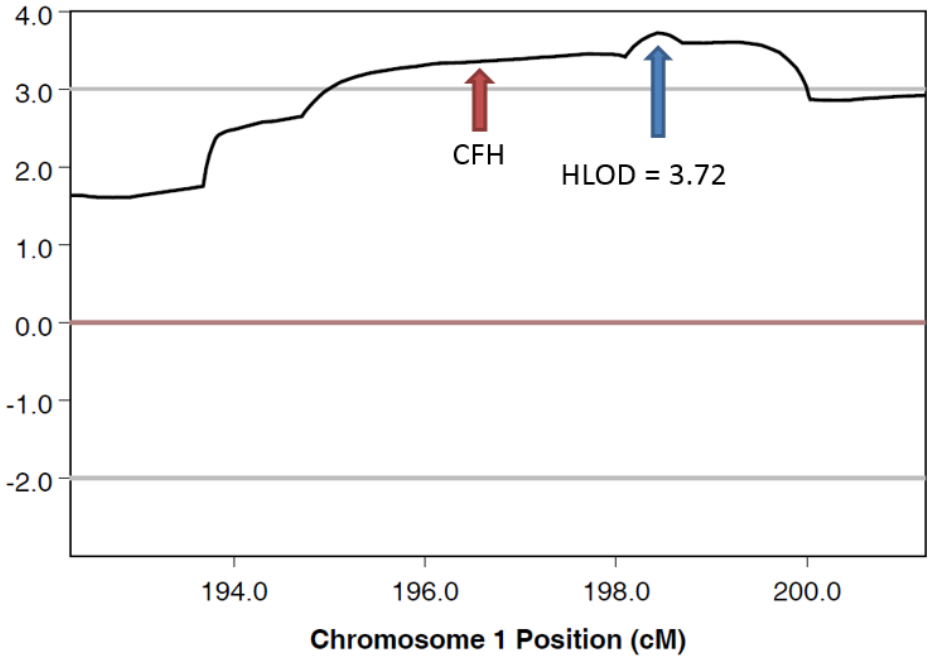
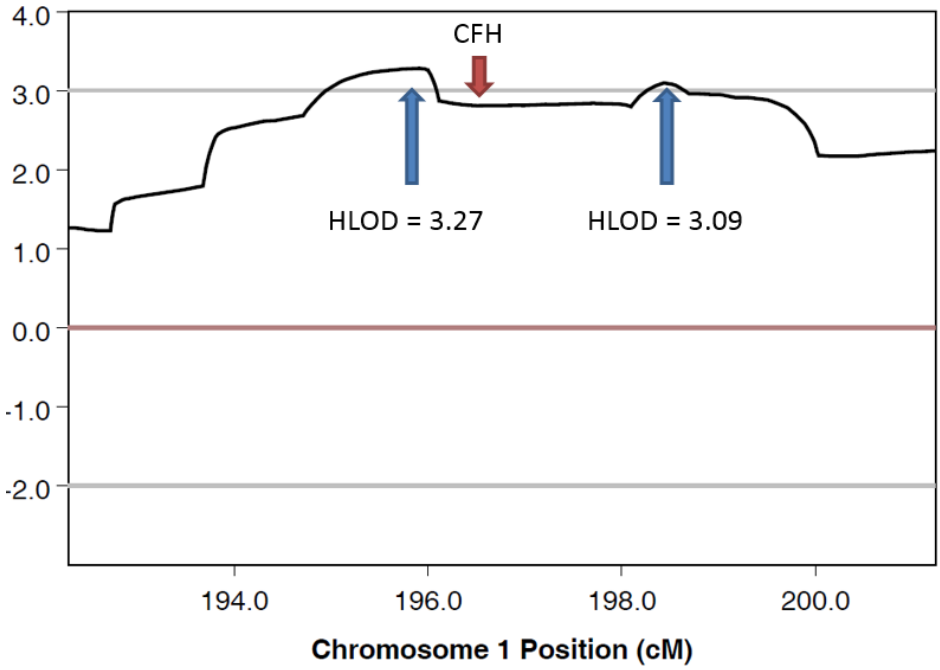


Figure 4.5 A-D: Multipoint linkage Results on Chromsome 1

Figure D. Incorporation of both the common and rare variants as covariates shows a reduction in the Maximum LOD score from 5.12 to 3.28.



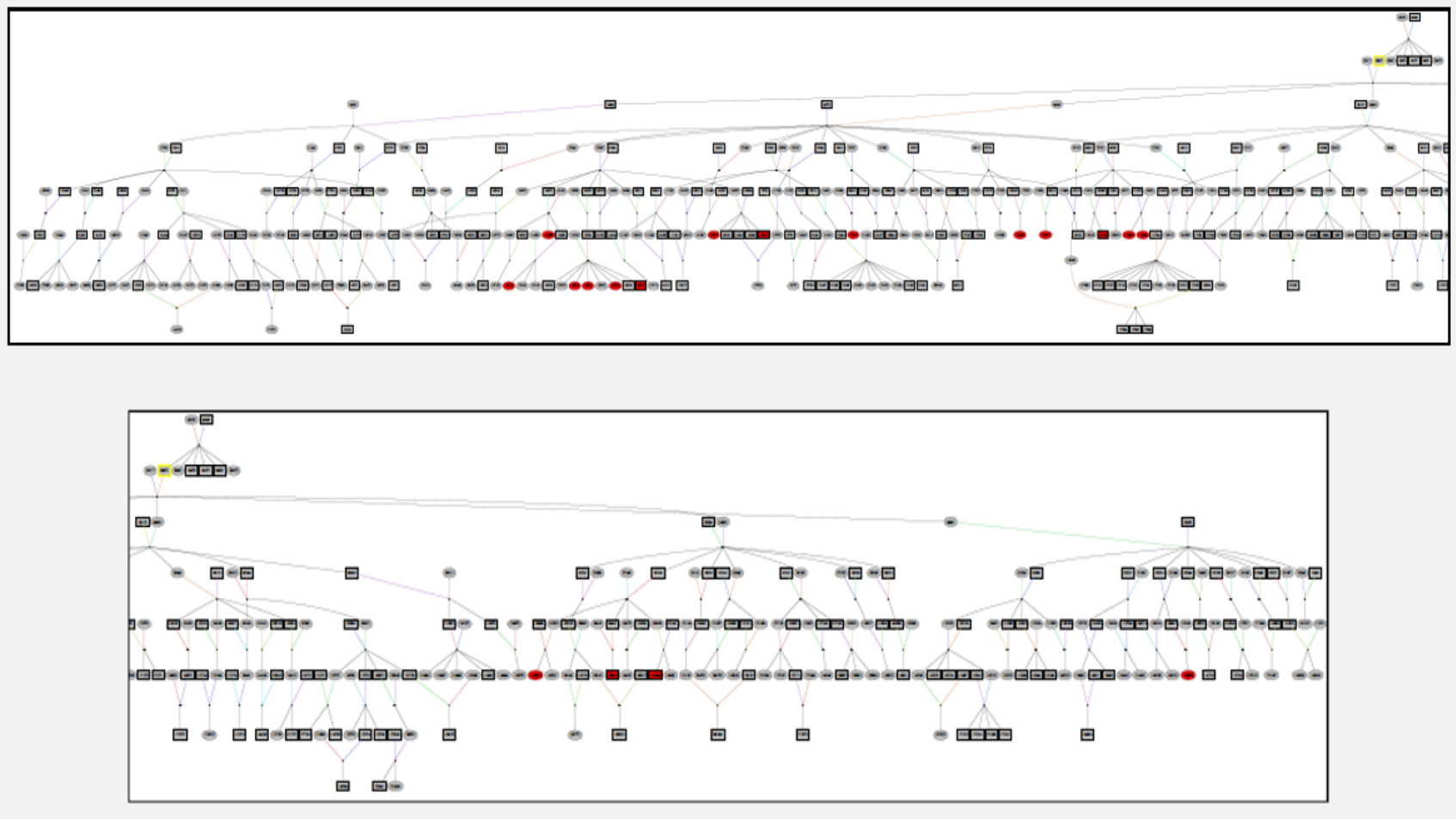
CFH is a regulator of the alternative complement cascade and associations to variation in *CFH* have consistently been replicated in AMD linkage and association analyses[46, 50, 55, 57, 134, 135]. *CFH* inhibits activation of complement component 3 (C3) to C3a and C3b, and in addition direct inactivation of C3b. Previous work has shown binding sites for C3b in short consensus repeat (SCR) domains1-4, 6-10, and 16-20 of complement factor H[151, 152]. The P503A variant is located within SCR domain 8, and thus may affect C3b binding affinity. Expanding the analysis of this variant to the full Amish dataset, we observe a total of 19 carriers including the four carriers in the nuclear Amish family, 11 reported as having AMD, 6 reported not having AMD, and 2 with an unknown affection status. Of the 19 self-reported carriers, 6 had their diagnosis confirmed by a retinal specialist.

Multipoint linkage analysis carried out across chromosome 1 shows an HLOD score above 5 within the region harboring the *CFH* rare and common Y402H variants. When including liability classes to account for these variants, we see a minimal reduction in the HLOD score, indicating that these variants may be only partially contributing to the observed linkage signal. Subsequent to the exome sequencing, 3 more members of the original nuclear family used to identify the P503A variant received a clinical diagnosis. Of these 3 subjects that originally self-reported as not having AMD, one individual, a-81, was seen by a retinal specialist and was reclassified as a case. The self-report age of exam (AOE) of this individual was 66 and the clinical AOE was 74, suggesting that this individual may have progressed to a non-severe case during the eight year time-span between the self-report and clinical exam. Genotyping results of a-81 showed that this person was a non-carrier of the P503A variant, but did carry one copy of the Y402H risk allele (Figure 4.2). AOE is a concern with diagnosing AMD due to the variable age of onset, where no control can definitively be classified without some chance of

progression to AMD. Our Amish study sample, although mainly classified through self-report status, is well represented with late age controls matched to cases, an important factor in phenotype assignment for AMD (Table 4.2). Of the other two unaffected siblings, a-83 was a carrier of both the P503A variant and A69S, and presented with small drusen at the time of exam. This reaffirms the need for further molecular characterization of this variant and its role in the AMD pathway.

Using the extensive genealogical data available in the AGDB, we observe that the carriers of this variant can be traced back four generations to a shared common ancestor and 15 of the 19 *CFH* rare variant carriers reside within Holmes County in Ohio (Figure 4.6). We fail to observe the *CFH* P503A variant in 2,247 non-Amish individuals (1,456 non-Amish cases and 791 non-Amish controls) or in publicly available databases.

Figure 4.6: Identification of the least common ancestor
Identifying the least common ancestor that is a carrier of the P503A variant illustrated by the subject highlighted in yellow.



This suggests that this *CFH* variant may have become enriched in the Ohio-Indiana Amish populations due to a recent founder event.

Acknowledgements:

We would like to thank the community members and family participants for agreeing to participate in this ongoing study. The research is supported by the National Institutes of Health grants AG019085, AG019726, EY21453-2, and AG044089. Some of the samples used in this study were collected while WKS and MAP-V were faculty members at Duke University. The authors would like to recognize the efforts of the late C. E. Jackson for his contributions to the overall Amish projects. We would also like to acknowledge work for this study that was performed using the Vanderbilt Center for Human Genetics Research Core facilities: the Genetic Studies Ascertainment Core, the DNA Resources Core, and the Computational Genomics Core.

Chapter V

CONCLUSIONS AND FUTURE DIRECTIONS

Age-related macular degeneration (AMD) is a debilitating, progressive disorder that has a large socioeconomic burden in the developed world with huge implications for current and future generations. As medical technology and healthcare standards continue to improve in both developed and developing countries, average life expectancy will increase. This growth will drive the incidence and prevalence of the disease to ever-increasing proportions. The increasing prevalence of AMD may be mitigated in the developed countries due to reduced rates of smoking, although this same trend is not currently seen in developing countries and it may become a growing factor for these countries. Understanding the genetic etiology of AMD and its relation to the multiple AMD phenotypes will have great implications for diagnosing and treating the disorder early enough in the disease process to preserve visual acuity.

As explained in chapter I, AMD shows substantial phenotypic heterogeneity throughout all stages of disease. Individuals affected with AMD may present with as little as intermediate drusen throughout their lives and not progress pass this stage, while other individuals present with combinations of geographic atrophy (GA), wet-AMD and drusen. Subjects with AMD also show heterogeneity with their rate of disease progression for both GA and neovascular AMD. Interestingly, for a complex disease, the genetic contribution to AMD is estimated to be very high, and in some cases as high as 70%[69]. This underlying genetic causality has led to great advances in the genetic study of AMD with linkage and association studies identifying replicable common variants that contribute upwards of 65% of the total genetic variation due to AMD.

We have also learned that genetic risk, when assessed in AMD is difficult to dissect with respect to the phenotypes, as the major genetic risk factors such as *CFH* and *ARMS2/HTRA1* trend in magnitude of effect depending on which of these phenotypes is being examined[153]. As an attempt to tease apart the phenotypic heterogeneity, we hypothesized that the major risk loci previously identified in genetic association analysis of AMD may in part explain drusen progression when treating drusen as a quantitative measure. To test this hypothesis it was necessary to take an imaging genetics approach using the major diagnostic measure, color fundus photographs, to extract drusen as a quantitative feature.

In chapter II we introduce the Age-related Eye Disease Study dataset due to its tremendous scientific value and heavy use in chapter 3, and discuss the image quality control procedures we employed for our quantitative analysis. The value of the AREDS data is significant as subjects were followed in a structured clinical trial for over 12 years with fundus photographic images taken bi-annually, making it an ideal dataset for a longitudinal study of drusen progression. The necessity of high quality fundus photos is of great importance to accurately quantify drusen load, mainly due to the phenotypic complexity of drusen in nature. Drusen can present as either soft or hard, leading to differences in size, number, color, border. As such, we examined the effects of image quality extensively in both the dbGAP based AREDS dataset and the CWRU/HIHG dataset. Based on manual review of digitized AREDS images, it was found that major quality issues were present and not well documented by dbGAP or AREDS.

We found that the distribution of high quality images is significantly different between the AREDS and the CWRU/HIHG data even after manual review. We believe that these auto generated quality score differences may be in part an artifact of the digitization process that took place with the AREDS data. Image quality after manual review did not have statistically significant impact on drusen calling and therefore we did

not take any additional steps with the quality information. Although this conversion process was a concern for our study, previous work examining the potential impact of the image digitization process has shown that major differences are not present when comparing parameters such as color, contrast of the original images to the digitized images[154]. Although this may be the case, we observed a significant quality drop within the AREDS data, and due to these quality issues we dropped roughly 30% of images as deemed unfit for analysis. After extensive quality control, we identified 973 images of sufficient quality to put forward for automated drusen quantification.

In chapter III we used the quality-controlled images to perform a quantitative genetic association analysis on drusen load. Under the assumption that the 19 common variants associated with AMD in the AMDGENE consortium study make up a bulk of the genetic variation contributing to AMD, we performed a hypothesis driven analysis examining these marker's contribution to drusen progression. We structured our analysis to include only the period under which subjects presented with drusen only, in attempt to isolate the role that these 19 loci contribute to drusen progression. To reduce heterogeneity in the dataset with respect to mean progression rates, we restricted our analysis to the first 6.5 years of follow-up data. Dropout of subjects due to advanced AMD progression led to an uneven representation of subjects at the end of the original 12.5-year study, driving the mean progression rate to reflect only subjects that finished the study with drusen only. As this was the case, we found that 6.5 years was the optimal follow-up time for a representative distribution of endpoint severity of drusen only, GA, and wet AMD.

We initially examined known non-genetic risk factors for their association with drusen progression. We did not observe sex, age, or smoking as contributing to drusen progression or baseline. It could be argued that age and smoking should show some contribution to drusen progression or baseline, as they are strong risk factors for AMD.

The absence of these associations in our results may be explained by the use of a case only approach in our analysis whereas these risk factors are typically enriched in case-control analyses.

Our primary genetic analysis was oriented toward using a genetic risk score to measure the cumulative effect of drusen load on baseline and progression. This allowed us to test a single hypothesis that cumulative genetic load contributes to drusen progression. Our results from the cumulative genetic risk score analysis showed a small contribution to baseline drusen load, but no association with drusen progression. It is important to note that the effect sizes used in the cumulative genetic risk score analysis were based on the results of the AMDGENE consortium which was assessing risk of disease, not progression of disease. In their analysis, the authors treated AMD phenotype status based on the presence or absence of disease, lumping intermediate AMD, geographic atrophy and wet AMD into one category. It is possible that these effect sizes do not effectively translate to just the intermediate phenotype of drusen only.

To test this, we also examined the 19 loci that make up the genetic risk score individually. Of these variants, the variant located at the *VEGFA* locus showed a nominal association with drusen progression in the AREDS dataset, although we were not able to replicate these results in the CWRU/HIHG dataset. VEGFA mRNA and protein are both enriched in drusen and are a factor in angiogenesis and inflammation. Again, this highlights the complicated interplay between the three phenotypic manifestations of AMD, with inflammation potentially relating to drusen biogenesis and angiogenesis relating to wet AMD. The major risk variant in *CFH* and *ARMS2* were not significantly associated with drusen progression or baseline. Interestingly, when we examine the intercepts for the *CFH* risk variant we see an apparent dominant pattern of effect in the AREDS data. This observation is also present in the CWRU/HIHG dataset, although the coefficients are not significant in either analysis. Under a dominant model,

we observe an allelic association with *CFH* when performing a Wilcoxon rank-sum test on drusen area at baseline in the AREDS sample (p=0.028). Again, we need to be critical of these results as this is not multiple test corrected.

The potential dominant effect of the *CFH* risk variant may be alluded to based on prior linkage analyses carried out by Klein et al. and Weeks et. al, which showed the *CFH* locus segregating in a dominant fashion with AMD[44, 155]. In the characterization of the AMD phenotype, the authors used soft drusen, GA, and neovascular AMD as their definitions of AMD. With the presence of multiple phenotypes for a case-control based diagnosis, it may be that the mode of inheritance is phenotype dependent. To further this analysis it would be ideal to test this in a larger sample. AREDS has recently made fundus photo data available on all subjects within the clinical trial on a single time point. This would give us the power to both see the possible dominant effect observed in our smaller datasets, and to see whether *CFH* truly contributes an allelic effect to baseline drusen.

To explore the role of unknown genetic variation to drusen progression, we used the data from the exome-chip based quantitative GWAS to perform a pathway analysis with the software package PARIS, using genes from the KEGG database. Our most highly significant pathway in the AREDS dataset was the cell-adhesion molecule pathway which is associated with inflammation (CAM; p-value < 0.0001). Through interrogation of the 127 genes that make up this pathway, we identify three genes driving the signal. These genes are *neurofascin (NFASC)* on chromosome 1, *CD226 molecule (CD226) on* chromosome 18, and *neurexin 1 (NRXN1)* on chromosome 2. CD226, which is the strongest associated driver gene in this analysis has been linked to regulation of proinflammatory responses and CD antigens have been reported as a drusen component[3, 156]. The statistical significance of these genes was confirmed using the

gene enrichment program “Versatile Gene-based Association Study” (VEGAS; Appendix D).

Future studies examining drusen progression are still needed. A critical observation that we made during the course of the study was the high level of intra-subject variation in the repeated measure data, and the inter-subject variation of subject’s baseline and progression rates. With drusen presenting with such complicated manifestations in color fundus photos, accurate quantification leaves room for improvement. It may be possible to reduce some of this variation by taking advantage of newer imaging technologies. One technology that is now being used for the clinical diagnosis of AMD is optical-coherence tomography (OCT). OCT allows the user to generate a 3-dimensional image of the RPE layer, allowing quantitative information to be collected on both drusen area and volume. The collection of this quantitative information may allow for better representation of the role of genetic variants in the natural history of drusen development. Current studies examining the correlation of OCT and fundus photography show reduced variation in OCT, although the identification of smaller drusen is stronger in color fundus photographs[157]. This observation may complicate matters even more as it again highlights the complexities of drusen analysis. We may find that it is necessary to integrate major imaging modalities such as OCT and color fundus photography together to accurately capture this AMD phenotype.

A natural observation in examining changes in drusen area over time are instances of spontaneous drusen regression. Drusen regression typically has been shown to precede the manifestation of GA and wet AMD. Although not significant in our analysis, we do observe a reduced slope in subjects that exit the study with wet AMD as compared to subjects that exit the study with GA or drusen only. Our AREDS data, although extensive, is limited by the time between visits that fundus photographs were taken and thus we do not have the resolution to characterize this phenomenon in our

samples. Although we had available more frequent visit data in the CWRU/HIHG set, the number of samples with this outcome data was limited.

Understanding the implications of drusen regression and its potential as an indicator for advancement to GA and wet AMD could benefit from quantitative drusen analysis. Some work has alluded to the role of drusen location and its potential effect on risk for AMD advancement[158]. These data were presented in the context of whether drusen found in the inner or outer rings of the grading grid play a role in progressing to advanced AMD[158]. By examining this spatial and temporal relationship between drusen and the advanced stages of disease, we may be able to better understand the mechanisms involved in drusen associated disease advancement at the site of insult.

To understand the genetic contribution of rare-variants to AMD, we examined the Amish populations of Ohio and Indiana in Chapter IV. Through exome sequencing of a highly penetrant nuclear Amish family we were able to identify a rare non-synonymous *CFH* variant (P503A) that is predicted to be damaging. We observed a significant genetic association of P503A when we expanded our analysis to include all subjects in our Amish sample. This association signal is complemented by a linkage scan across the *CFH* locus, which shows that the P503A variant has some contribution to the overall linkage peak. This variant was not observed in a cohort of 1,400 non-Amish subjects. We recently genotyped 700 subjects from the Pennsylvania Amish community and do not observe the P503A variant in this population either, leading us to surmise that this is a founder mutation specific to the Ohio and Indiana Amish. We must note we do not have phenotype data on the 700 Pennsylvania Amish subjects and thus this distribution of cases and controls is unknown.

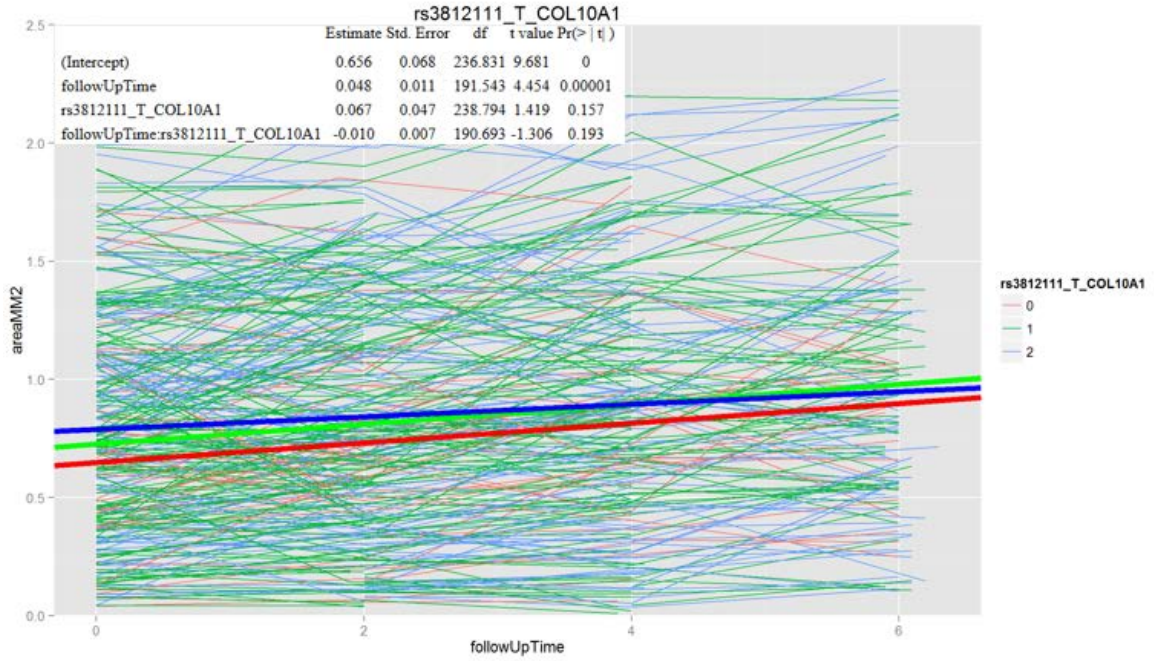
Important to the future study of our Amish cohort for AMD is better characterization of the phenotype data. Our current dataset relies on self-report diagnostic information to describe AMD case status. The results of our self-report study

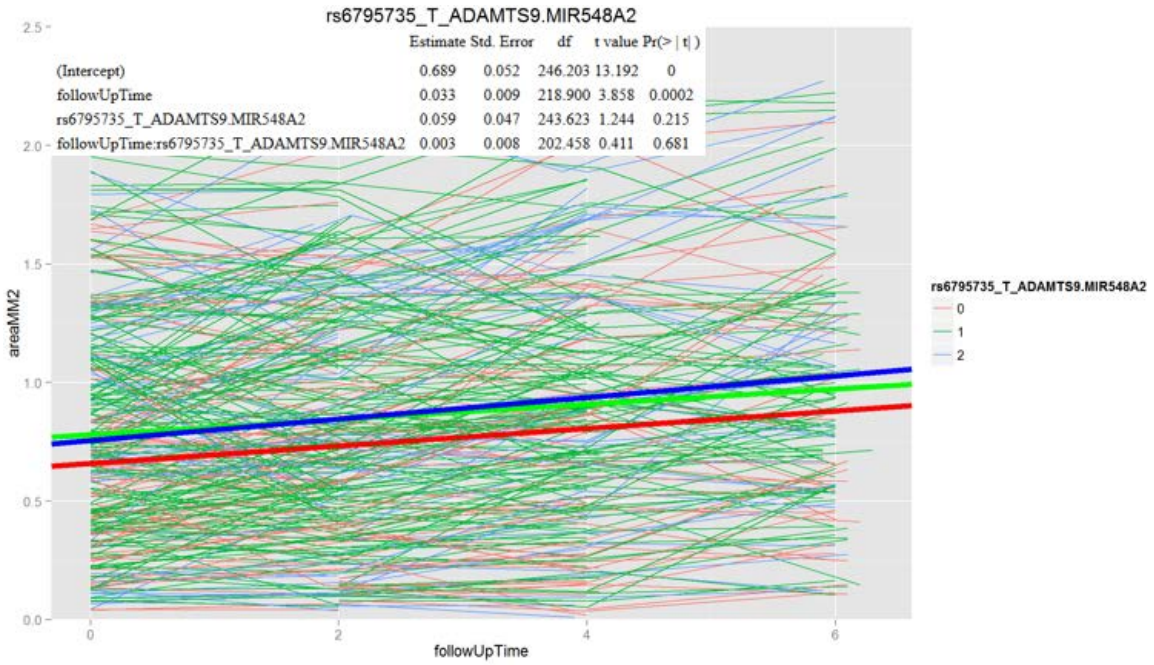
were promising in that this measure may serve a purpose, as a surrogate for the assignment of case-control status for our AMD related studies, but true diagnoses by a retinal specialist is a necessity considering the phenotypic complexities of the disease. Also of importance to the Amish AMD studies is discovering the missing genetic contribution to AMD. Our risk score analysis showed reduced genetic risk based on the known 19 common variants, and the rare *CFH* variant only appears to explain some of that unknown genetic risk. Understanding the unknown common and rare genetic burden of disease will need to be further explored.

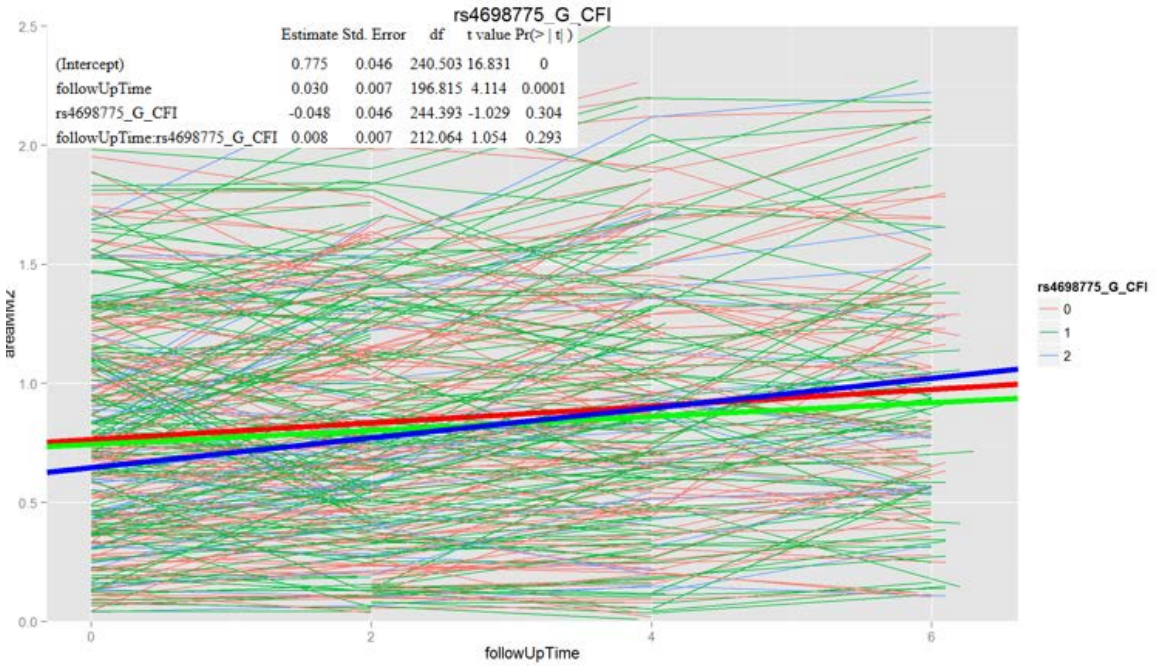
In conclusion, AMD continues to be a valuable complex disease for genetic studies due to its high heritability and its socioeconomic impact. Although a large proportion of the genetic variation explained by AMD has been identified, its contribution is largely put in the context of disease risk. Many of the current studies examining progression are focused on binary outcomes trying to understand the impact of progressing from early to late AMD with little effort put towards understanding rate of progression. Elucidation of potential rare and common genetic contributors for AMD rate of progression may lead to better understanding of the mechanisms involved in AMD pathogenesis and strides in AMD clinical management.

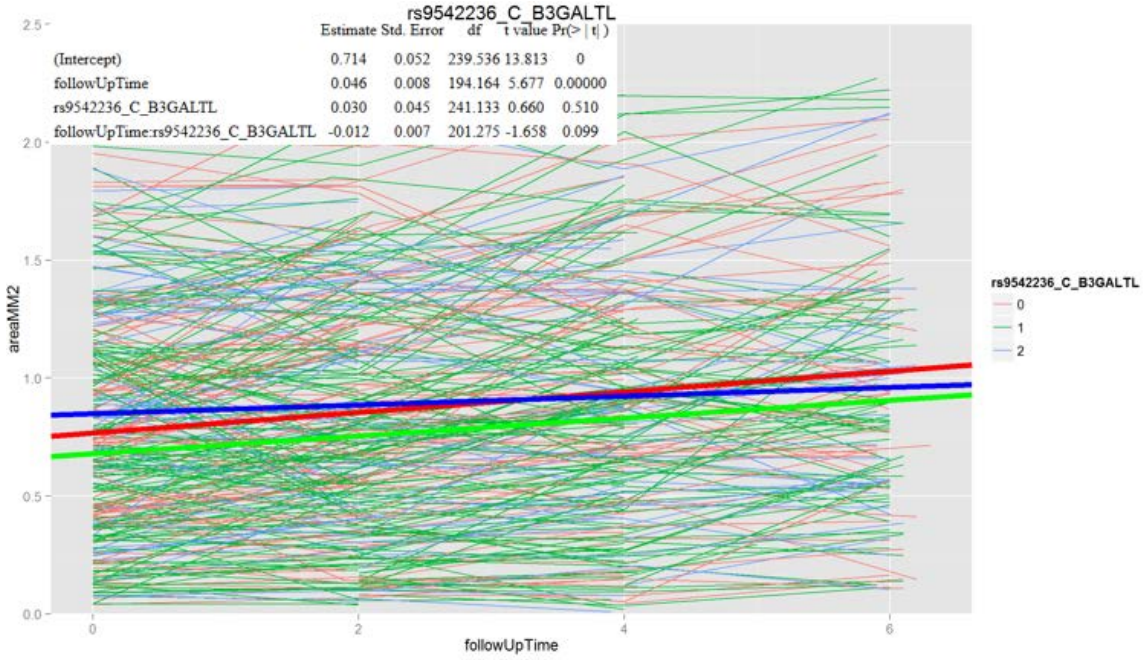
APPENDIX

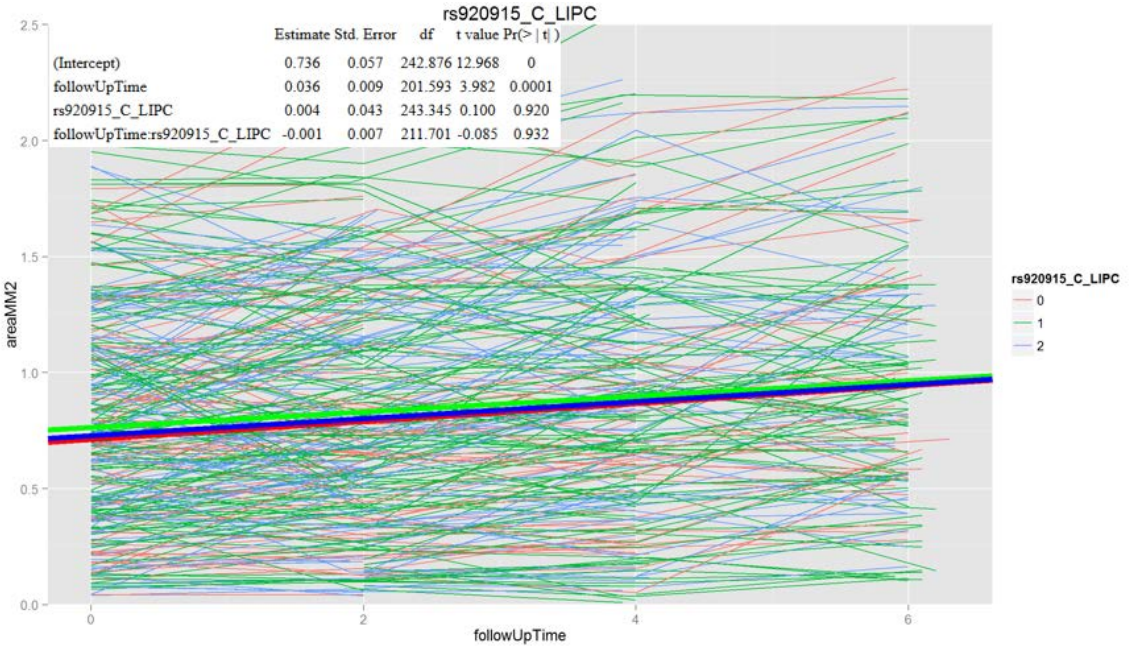
Appendix A: Progression Slopes for 15 Additional Contributors to the Cumulative Genetic Risk Score: Represented by the AREDS dataset

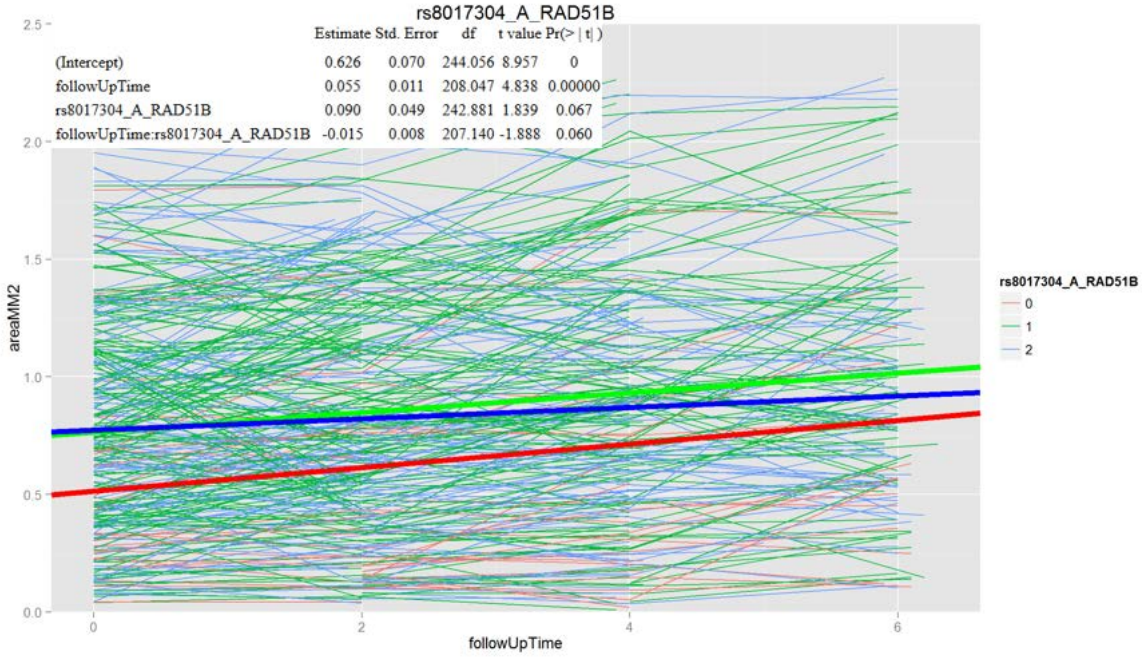


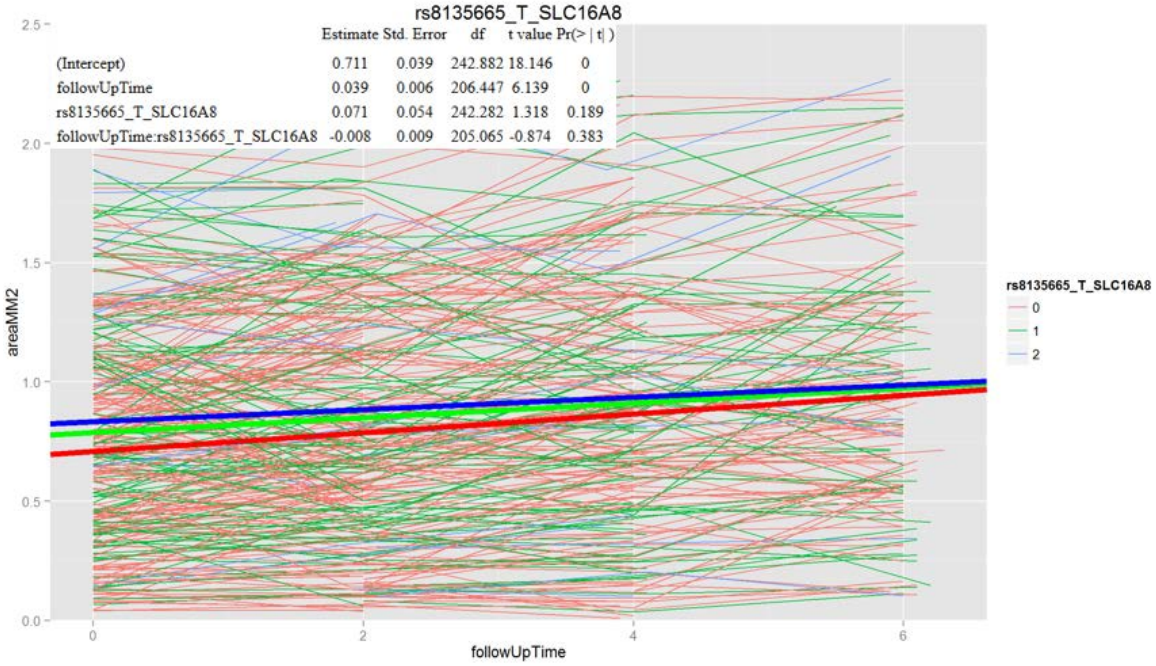


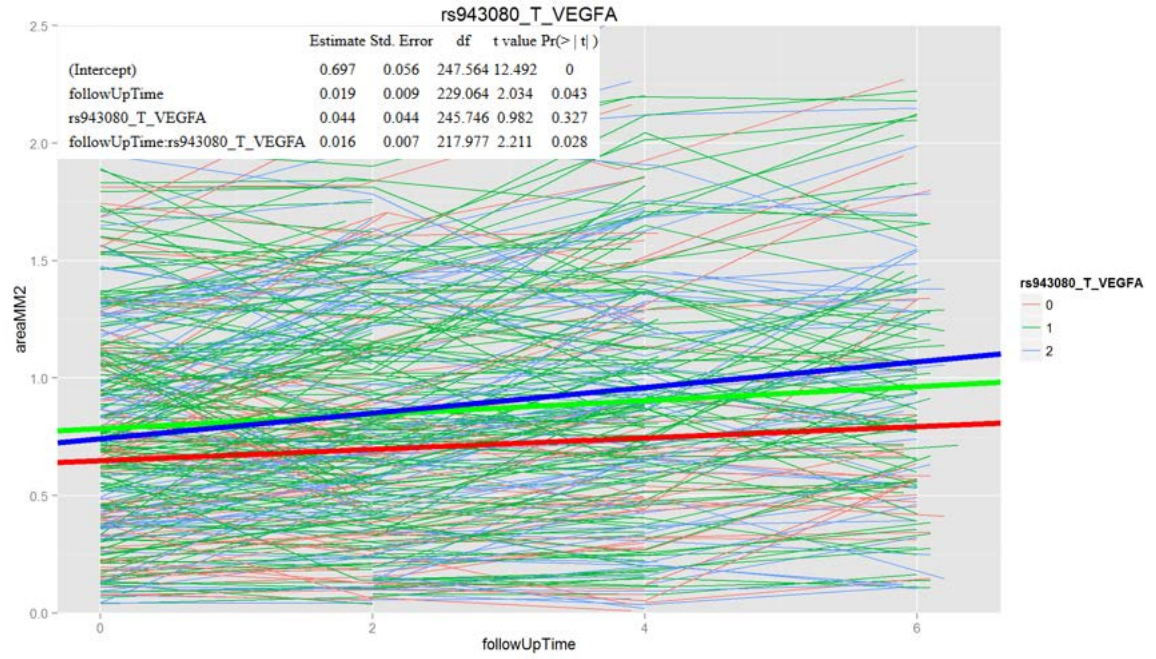

rs4698775_G_CFI
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.775 0.046 240.503 16.831 0
followUpTime 0.030 0.007 196.815 4.114 0.0001
rs4698775_G_CFI -0.048 0.046 244.393 -1.029 0.304
followUpTime:rs4698775_G_CFI 0.008 0.007 212.064 1.054 0.293
areaMM2
followUpTime
rs4698775_G_CFI
0
1
2

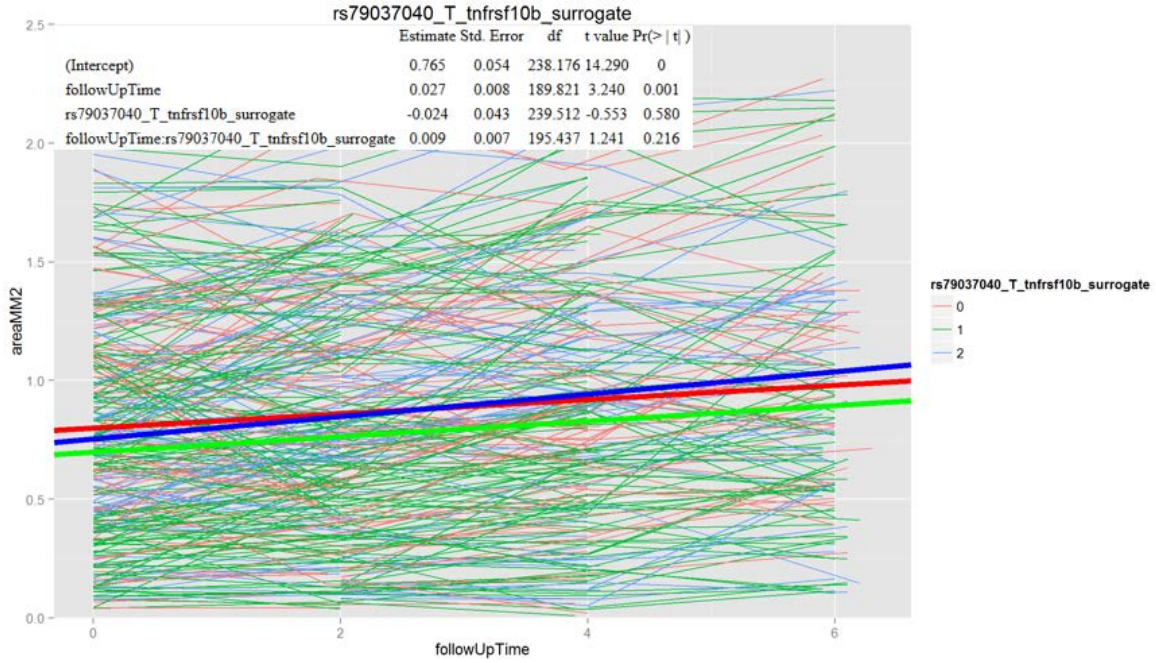

rs9542236_C_B3GALTL
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.714 0.052 239.536 13.813 0
followUpTime 0.046 0.008 194.164 5.677 0.00000
rs9542236_C_B3GALTL 0.030 0.045 241.133 0.660 0.510
followUpTime:rs9542236_C_B3GALTL -0.012 0.007 201.275 -1.658 0.099
areaMM2
followUpTime
rs9542236_C_B3GALTL
0
1
2

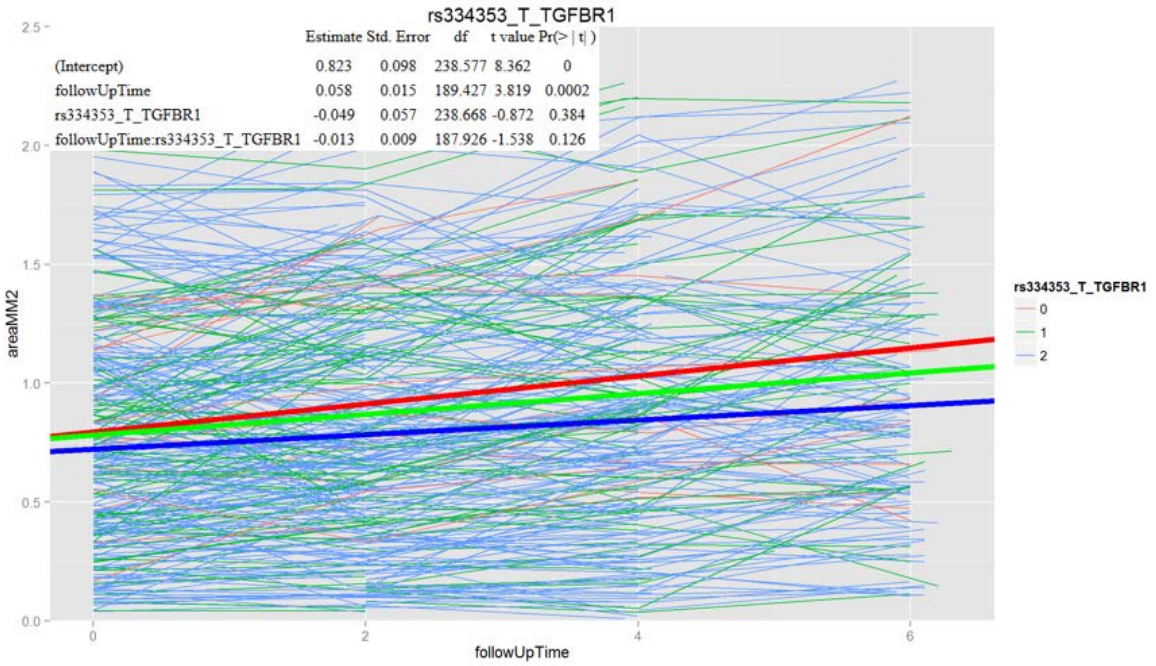

rs920915_C_LIPC
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.736 0.057 242.876 12.968 0
followUpTime 0.036 0.009 201.593 3.982 0.0001
rs920915_C_LIPC 0.004 0.043 243.345 0.100 0.920
followUpTime:rs920915_C_LIPC -0.001 0.007 211.701 -0.085 0.932
areaMM2
followUpTime
rs920915_C_LIPC
0
1
2

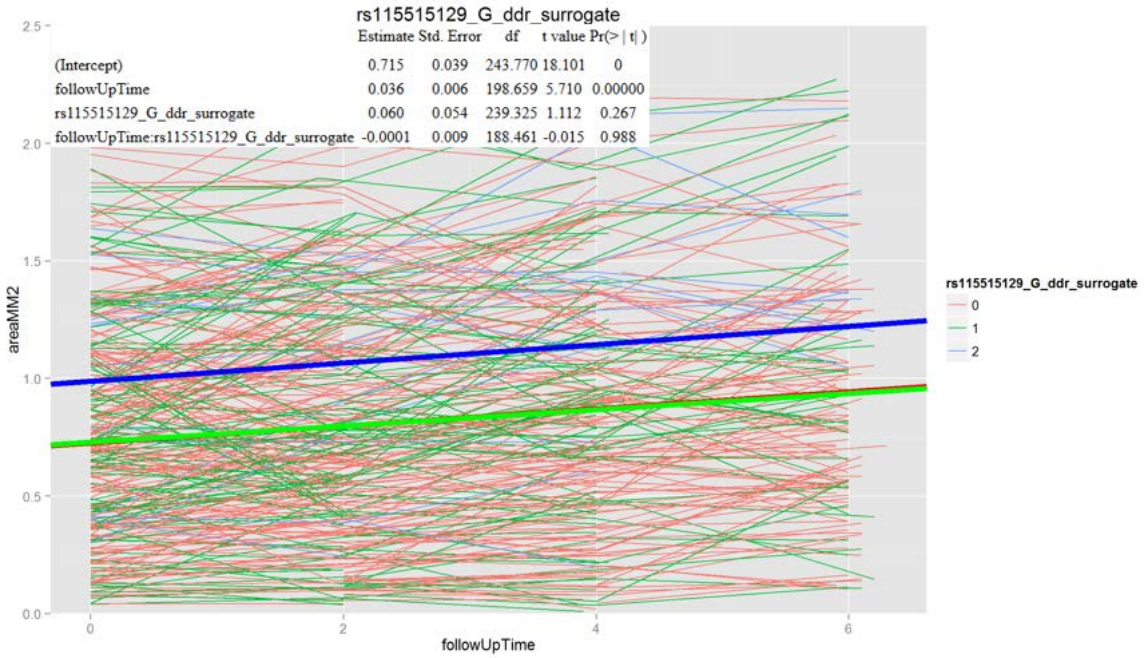

rs8017304_A_RAD51B
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.626 0.070 244.056 8.957 0
followUpTime 0.055 0.011 208.047 4.838 0.00000
rs8017304_A_RAD51B 0.090 0.049 242.881 1.839 0.067
followUpTime:rs8017304_A_RAD51B -0.015 0.008 207.140 -1.888 0.060
areaMM2
followUpTime
rs8017304_A_RAD51B
0
1
2

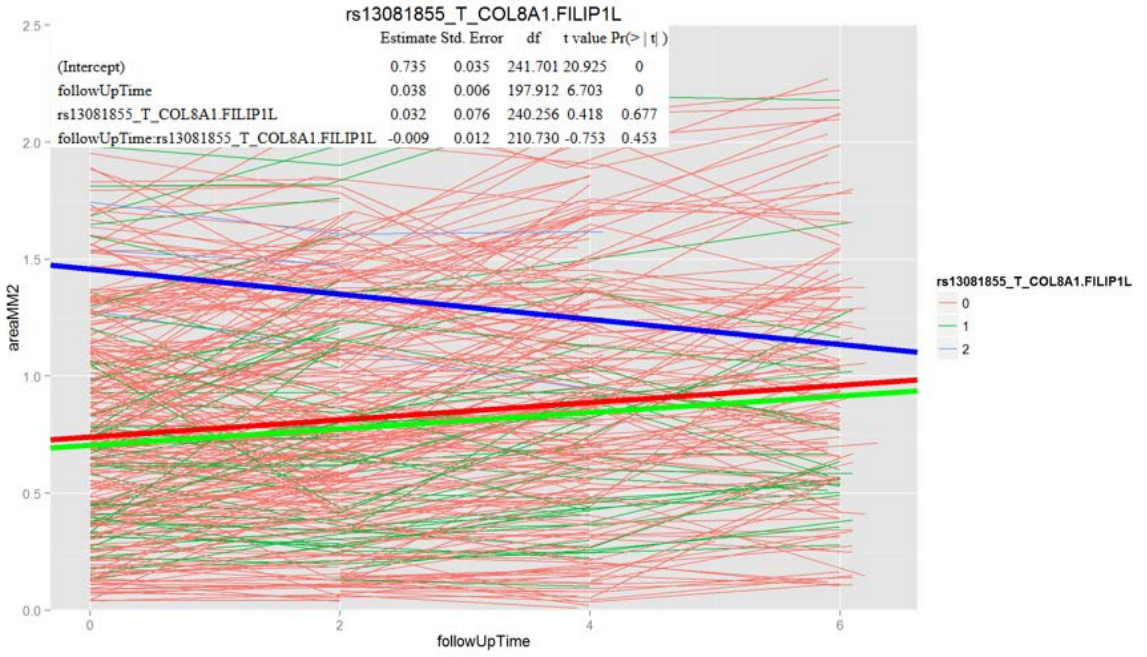

rs8135665_T_SLC16A8
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.711 0.039 242.882 18.146 0
followUpTime 0.039 0.006 206.447 6.139 0
rs8135665_T_SLC16A8 0.071 0.054 242.282 1.318 0.189
followUpTime:rs8135665_T_SLC16A8 -0.008 0.009 205.065 -0.874 0.383
areaMM2
followUpTime
rs8135665_T_SLC16A8
0
1
2

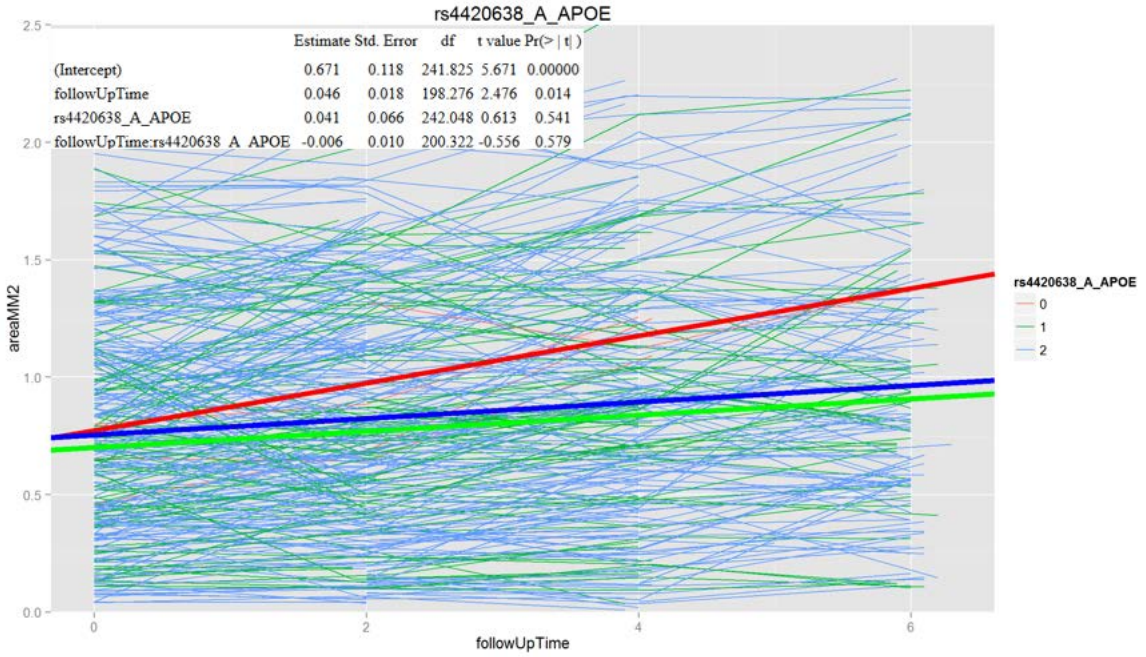

rs943080_T_VEGFA
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.697 0.056 247.564 12.492 0
followUpTime 0.019 0.009 229.064 2.034 0.043
rs943080_T_VEGFA 0.044 0.044 245.746 0.982 0.327
followUpTime:rs943080_T_VEGFA 0.016 0.007 217.977 2.211 0.028
areaMM2
followUpTime
rs943080_T_VEGFA
0
1
2

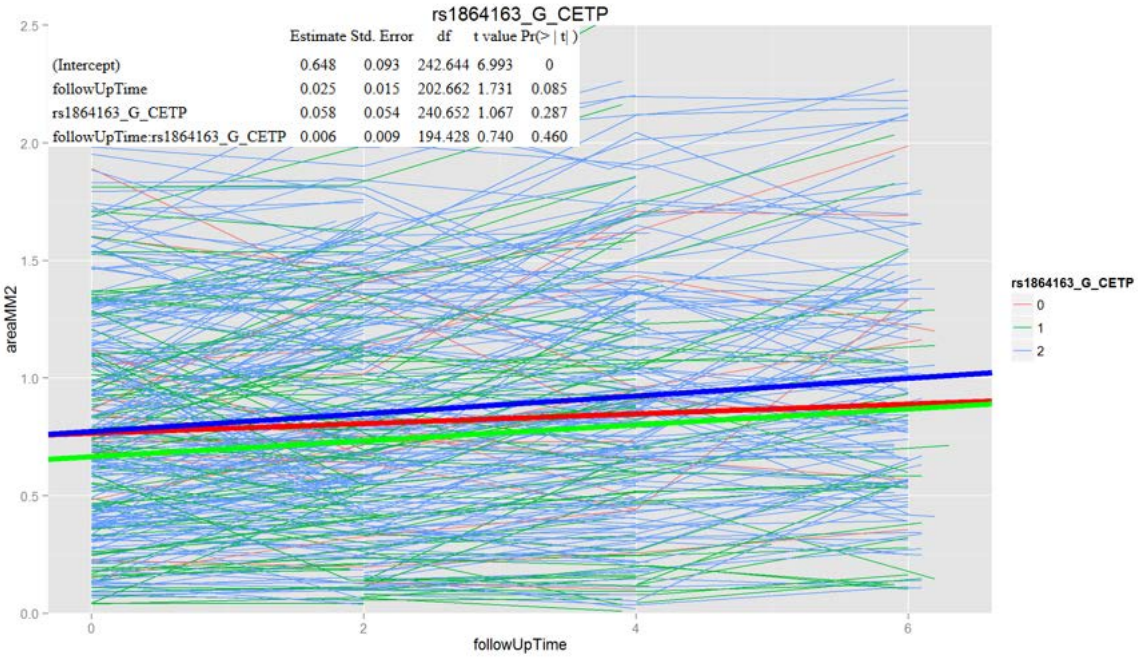

rs79037040_T_tnfrsf10b_surrogate
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.765 0.054 238.176 14.290 0
followUpTime 0.027 0.008 189.821 3.240 0.001
rs79037040_T_tnfrsf10b_surrogate -0.024 0.043 239.512 -0.553 0.580
followUpTime:rs79037040_T_tnfrsf10b_surrogate 0.009 0.007 195.437 1.241 0.216
areaMM2
followUpTime
rs79037040_T_tnfrsf10b_surrogate
0
1
2

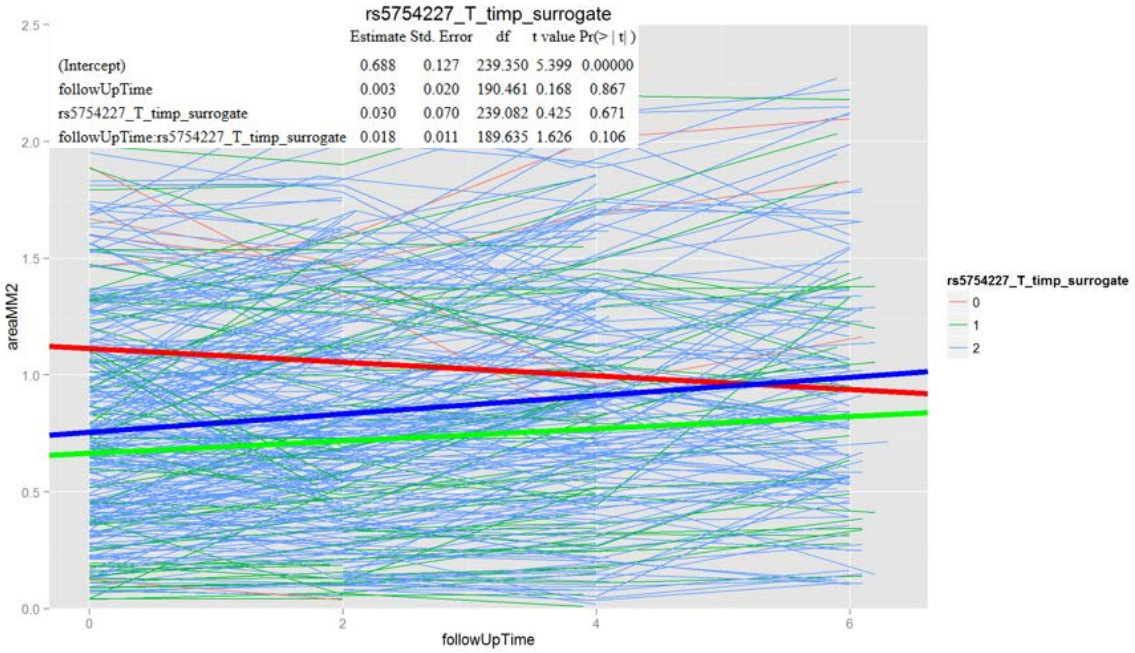

rs334353_T_TGFBR1
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.823 0.098 238.577 8.362 0
followUpTime 0.058 0.015 189.427 3.819 0.0002
rs334353_T_TGFBR1 -0.049 0.057 238.668 -0.872 0.384
followUpTime:rs334353_T_TGFBR1 -0.013 0.009 187.926 -1.538 0.126
areaMM2
followUpTime
rs334353_T_TGFBR1
0
1
2


rs115515129_G_ddr_surrogate
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.715 0.039 243.770 18.101 0
followUpTime 0.036 0.006 198.659 5.710 0.00000
rs115515129_G_ddr_surrogate 0.060 0.054 239.325 1.112 0.267
followUpTime:rs115515129_G_ddr_surrogate -0.0001 0.009 188.461 -0.015 0.988
areaMM2
followUpTime
rs115515129_G_ddr_surrogate
0
1
2


rs13081855_T_COL8A1.FILIP1L
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.735 0.035 241.701 20.925 0
followUpTime 0.038 0.006 197.912 6.703 0
rs13081855_T_COL8A1.FILIP1L 0.032 0.076 240.256 0.418 0.677
followUpTime:rs13081855_T_COL8A1.FILIP1L -0.009 0.012 210.730 -0.753 0.453
areaMM2
followUpTime
rs13081855_T_COL8A1.FILIP1L
0
1
2


rs4420638_A_APOE
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.671 0.118 241.825 5.671 0.00000
followUpTime 0.046 0.018 198.276 2.476 0.014
rs4420638_A_APOE 0.041 0.066 242.048 0.613 0.541
followUpTime:rs4420638_A_APOE -0.006 0.010 200.322 -0.556 0.579
areaMM2
followUpTime
rs4420638_A_APOE
0
1
2


rs1864163_G_CETP
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.648 0.093 242.644 6.993 0
followUpTime 0.025 0.015 202.662 1.731 0.085
rs1864163_G_CETP 0.058 0.054 240.652 1.067 0.287
followUpTime:rs1864163_G_CETP 0.006 0.009 194.428 0.740 0.460
areaMM2
followUpTime
rs1864163_G_CETP
0
1
2


rs5754227_T_timp_surrogate
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 0.688 0.127 239.350 5.399 0.00000
followUpTime 0.003 0.020 190.461 0.168 0.867
rs5754227_T_timp_surrogate 0.030 0.070 239.082 0.425 0.671
followUpTime:rs5754227_T_timp_surrogate 0.018 0.011 189.635 1.626 0.106
areaMM2
followUpTime
rs5754227_T_timp_surrogate
0
1
2

Appendix B: Tabulated Results of Top Progression Quantitative GWAS Hits

CHR	SNP	BP	BETA AREDS	P-value AREDS	P-value CWRU / Miami
1	rs2069084	234984988	-0.02518	1.52E-06	0.2582
2	exm279714	238668783	-0.03331	6.16E-06	0.5632
7	rs10226466	96357756	0.02789	1.27E-05	0.269
17	rs8077882	72145172	0.01507	1.37E-05	0.8151
6	rs6568924	116596247	-0.01717	1.66E-05	0.6334
2	rs10203272	86033879	-0.02126	1.71E-05	0.9214
6	rs6911639	32978178	0.0181	2.33E-05	0.2415
5	rs425203	57917792	-0.0154	2.39E-05	0.9321
5	rs4304068	158976183	-0.01736	2.42E-05	0.2061
7	rs395158	9216593	0.0179	4.01E-05	0.4838
8	rs1546745	90545792	0.01459	4.44E-05	0.3766
17	rs7214008	72148066	-0.01503	4.77E-05	0.8671
8	rs7846085	90528845	0.01457	4.80E-05	0.3766
8	rs11786321	90409007	0.01424	4.83E-05	0.2956
8	rs12544520	90407910	-0.01721	4.91E-05	0.26
1	rs12737855	204912843	-0.01653	4.93E-05	0.1352
4	rs2011590	38944101	0.01425	5.20E-05	0.1997
1	rs2796160	234966277	-0.02238	6.60E-05	0.2933
8	rs10956094	90478557	0.0141	6.77E-05	0.3588
5	rs1428609	66680616	0.01369	6.89E-05	0.3383
8	rs1386648	90539951	0.01428	7.25E-05	0.299
6	exm573430	116575116	0.02026	8.51E-05	0.9273
20	rs6127826	55307132	0.02811	8.60E-05	0.7405
8	rs1510465	90576635	0.01449	8.72E-05	0.3571
13	rs9582867	105099441	0.01696	0.0001048	0.04434
14	rs2236135	23595721	-0.01813	0.0001204	0.9356
13	rs111689935	31787926	0.02324	0.0001213	0.499
8	rs9297680	90597465	-0.01393	0.0001257	0.4882
16	rs11861081	8321930	0.01555	0.0001276	0.6308
13	rs9552533	22454710	-0.02904	0.0001465	0.1968
7	exm2270701	137089307	-0.01356	0.0001477	0.01182
18	rs7228563	71408712	0.01327	0.0001595	0.004997
6	rs62414126	116460264	0.02534	0.0001631	0.6724
7	rs3919487	144955976	-0.01428	0.0001867	0.2857
6	rs62414146	116561087	0.02476	0.0001993	0.6724
22	rs9606708	30646126	0.01401	0.0001996	N/A
1	rs11590192	231922845	-0.01296	0.0002128	0.688
13	rs4641600	110500338	-0.01506	0.0002156	0.06816
2	rs3943477	5224268	0.0159	0.0002196	0.1723

18	rs12956584	72348611	0.01772	0.0002245	0.9689
3	kgp11084897	128562396	0.02877	0.0002341	0.5946
3	rs234044	172313268	-0.0132	0.0002419	0.5018
6	rs850577	97365239	0.01467	0.0002424	0.09543
15	rs7178749	62053897	-0.01326	0.0002528	0.1125
12	rs427244	124605287	0.01663	0.0002583	0.1506
17	rs11652016	49570481	-0.01308	0.0002601	0.1479
1	rs17014713	111763153	0.02159	0.0002635	0.6172
8	rs6981722	138769527	0.01362	0.0002641	0.3081
2	rs9288260	197127830	0.02196	0.0002642	0.8317
14	rs10138170	22279576	0.01438	0.0002658	0.9963
1	rs10923507	118731761	-0.01418	0.0002667	0.6933
2	rs6729441	217048565	0.02185	0.0002672	0.8764
2	rs7583902	217057937	0.02185	0.0002672	0.8764
2	rs1051685	217070376	0.02185	0.0002672	0.8764
10	rs2590289	60059641	-0.01952	0.000269	0.7631
21	rs9978712	17820112	0.01364	0.0002711	0.4424
19	rs17216041	19366643	-0.01549	0.0002855	0.579
2	rs13407838	113745428	-0.01763	0.0002875	0.4538
1	rs17257729	118678953	-0.01311	0.0002905	0.9516
18	rs1942467	70517206	-0.0249	0.0002906	0.684
10	rs7895100	87683328	0.01725	0.0002926	0.7544
13	rs11620257	31776738	0.02208	0.0002965	0.1885
4	rs13115030	87296065	-0.01634	0.0002975	0.1542
6	rs12527153	116491302	0.02386	0.0002988	0.7087
4	rs16891600	15210390	-0.02538	0.0003037	0.6155
5	rs6450023	68526063	0.01457	0.0003099	0.378
15	rs7162855	58811252	-0.02107	0.00031	0.9917
18	rs1474128	24992045	0.0126	0.0003168	0.9928
20	rs6060989	30429763	-0.01568	0.0003181	0.5371
4	rs17751557	87834607	0.01626	0.0003228	0.8545
12	rs7138639	98538264	0.01271	0.0003228	0.2455
1	rs4411121	118757034	-0.01393	0.0003266	0.892
4	rs11097109	87269343	-0.01617	0.000327	0.1861
9	rs11142863	74135468	-0.01849	0.0003278	0.6431
4	rs10050311	87754419	0.01756	0.0003336	0.295
13	rs9513993	102861383	-0.02876	0.0003357	0.8503
1	rs4325188	105700935	-0.01396	0.0003359	0.1898
1	rs6427122	168274691	0.01333	0.0003403	0.03495
8	rs11780016	90601100	0.015	0.000343	0.7974
3	rs9860340	87783976	0.0158	0.0003449	0.1326

10	rs946516	80609135	-0.01367	0.0003454	0.6136
19	rs1064395	19361735	-0.01523	0.0003464	0.579
11	rs10837204	39671205	0.01472	0.0003519	0.9432
5	rs11134701	170972663	0.01995	0.0003526	0.6288
18	rs1144081	50087903	0.01435	0.0003638	0.002056
17	rs4792394	13684917	-0.01253	0.0003666	0.5361
10	rs6482124	18298132	-0.01602	0.0003768	0.1063
8	rs13282836	26678206	-0.01568	0.0003821	0.3995
2	rs1453780	5226989	0.01575	0.0003853	0.1755
6	rs883273	154630541	-0.0133	0.0003996	0.9036
5	rs7717348	95407894	0.01402	0.0004097	0.7824
11	rs518119	124181158	0.01855	0.0004125	0.5714
10	rs7073090	59943234	-0.01982	0.0004196	0.6953
6	rs6912724	144814016	-0.01395	0.0004217	0.01083
18	rs1144076	50079024	0.0147	0.000425	0.002048
8	rs7825497	714648	0.01757	0.0004391	0.3271
4	rs6849805	84626069	-0.0159	0.0004406	0.3539
7	rs7776689	96162825	0.01485	0.0004504	0.904
2	rs4674761	224071098	-0.01586	0.0004671	0.7444
11	rs12788923	37083616	-0.01303	0.0004681	0.9554
7	rs6963647	96170054	0.01537	0.0004707	0.9852
11	rs1503932	11206934	0.01251	0.0004772	0.8493
9	rs12375980	11178672	-0.01255	0.00049	0.7574
10	rs7900951	133657703	-0.0135	0.0004932	0.1942
4	rs6819155	87976178	0.01537	0.0004941	0.9198
5	rs10475598	173648438	-0.01317	0.0004976	0.6329

Appendix C: Tabulated Results of Top Pathways from Paris

Pathway results are based on AREDS quantitative GWAS.

Pathway ID	Pathway Name	Total SNP Count	Description	P-value AREDS
197335	hsa04514	2277	Cell adhesion molecules (CAMs)	0.0001
197353	hsa04670	1686	Leukocyte transendothelial migration	0.00055
197306	hsa03450	127	Non-homologous end-joining	0.0008
197277	hsa00780	37	Biotin metabolism	0.0011
197268	hsa00640	392	Propanoate metabolism	0.0016
197361	hsa04810	2502	Regulation of actin cytoskeleton	0.005
197333	hsa04510	2848	Focal adhesion	0.00655
197343	hsa04621	484	NOD-like receptor signaling pathway	0.00685
197379	hsa05210	1034	Colorectal cancer	0.0078
197292	hsa02010	649	ABC transporters	0.00875
197381	hsa05212	782	Pancreatic cancer	0.00885
197378	hsa05200	4423	Pathways in cancer	0.0094
197314	hsa04110	1009	Cell cycle	0.0138
197209	hsa00052	321	Galactose metabolism	0.0161
197283	hsa00903	143	Limonene and pinene degradation	0.01805
197391	hsa05222	1272	Small cell lung cancer	0.0194
197389	hsa05220	837	Chronic myeloid leukemia	0.02695
197279	hsa00790	105	Folate biosynthesis	0.0278
197336	hsa04520	1446	Adherens junction	0.0285
197269	hsa00650	347	Butanoate metabolism	0.02895
197350	hsa04662	943	B cell receptor signaling pathway	0.03055
197380	hsa05211	773	Renal cell carcinoma	0.0325
197227	hsa00280	485	Valine, leucine and isoleucine degradation	0.0335
197265	hsa00604	260	Glycosphingolipid biosynthesis - ganglio series	0.0379
197344	hsa04622	459	RIG-I-like receptor signaling pathway	0.0385
197331	hsa04360	2345	Axon guidance	0.04635
197346	hsa04630	1224	Jak-STAT signaling pathway	0.0466

Appendix D: Tabulated Results of Gene Enrichment Analysis (VEGAS)
Table is based on AREDS quantitative GWAS results.

CHR	Gene	# SNPs	Start	Stop	P-value
6	BRD2	16	33044414	33057260	2.00E-04
20	DUSP15	9	29912530	29922140	0.000363
6	HLA-DOA	20	33079937	33085367	0.000449
22	NAGA	1	40784283	40796792	0.000524
22	FAM109B	3	40800200	40805388	0.000674
22	C22orf32	3	40805644	40810234	0.000677
22	NDUFA6	3	40811475	40816834	0.000681
20	C20orf3	7	24891579	24921425	0.000725
19	TM6SF2	7	19236173	19245074	0.000742
20	C20orf3	7	24891579	24921425	0.000744
19	HAPLN4	7	19227470	19234560	0.000761
20	TPX2	5	29790564	29853264	0.000773
17	TBC1D3E-1	1	33358341	33369482	0.000817
17	TBC1D3-1	1	33358400	33369299	0.000834
17	TBC1D3E-1	1	33358341	33369482	0.000844
17	TBC1D3F-1	1	33358341	33369482	0.000867
19	NCAN	11	19183781	19224061	0.000874
22	CYP2D6	6	40852444	40856827	0.001008
20	CST7	7	24877865	24888562	0.001022
16	SHCBP1	1	45171968	45212812	0.001034
1	SPAG17	12	118297810	118529357	0.001073
16	VPS35	3	45251089	45280645	0.001095
19	SF4	8	19248321	19292307	0.001159
7	SHFM1	7	96156014	96177139	0.00116
20	FOXS1	8	29895763	29897081	0.001162
22	C1QTNF6	18	35906151	35914276	0.00146
22	TCF20	8	40885962	40941389	0.001509
10	IPMK	5	59625619	59697700	0.00155
21	SIM2	13	36993860	37044380	0.001567
18	CD226	8	65681172	65775212	0.001568
20	MYLK2	7	29870838	29886161	0.001591
1	F11R	12	159231624	159257757	0.001726
7	FLJ43692	2	143514609	143523669	0.001824
11	METT5D1	6	28086373	28311630	0.00185
22	WBP2NL	6	40724737	40754423	0.00209
1	CTNNBIP1	3	9830920	9892903	0.00214
7	CTAGE4-1	2	143511480	143514106	0.00219
11	KIF18A	5	27998738	28086322	0.002272

20	BCL2L1	6	29715921	29774317	0.00232
6	TSPYL4	14	116677823	116681954	0.00237
22	GSTT1	2	22706138	22714284	0.002373
4	PTPN13	9	87734908	87955326	0.00241
6	TSPYL1	14	116704438	116707973	0.002444
21	KRTAP20-2	5	30929453	30929651	0.00245
11	OR8G1	7	123625632	123640966	0.00248
6	DSE	26	116707975	116866135	0.00253
20	TTLL9	12	29922165	29994519	0.00254
14	SIX4	2	60246008	60260545	0.002559
11	OR8G5	7	123639932	123640973	0.00259
1	RBBP5	11	203322601	203357754	0.00262
4	C4orf36	9	88016381	88032599	0.00263
18	WDR7	32	52469613	52848034	0.00264
21	KRTAP20-1	5	30910644	30910815	0.00264
1	ITLN2	11	159181439	159191213	0.0027
7	RARRES2	9	149666350	149669639	0.00282
12	NEDD1	8	95825374	95870172	0.00287
7	C7orf29	9	149657870	149660743	0.00288
7	LRRC61	10	149651537	149666172	0.00288
22	LOC652968	6	29011106	29015616	0.00313
4	ATOH1	12	94969100	94970165	0.00314
7	REPIN1	10	149696811	149702066	0.003148
10	REEP3	10	64951128	65051978	0.003183
14	SGPP1	1	63220687	63264509	0.00319
22	TBC1D10A	8	29017978	29052894	0.003259
21	KRTAP20-3	4	30937053	30937326	0.003387
22	OSM	6	28988818	28992840	0.0034
8	FAM164A	6	79740884	79792490	0.003615
19	GATAD2A	5	19357641	19480741	0.00383
19	RASGRP4	5	43591537	43608785	0.00387
22	C22orf26	5	44825002	44828688	0.00389
14	MNAT1	6	60271222	60505151	0.00391
6	C6orf154	3	43582684	43586402	0.00392
12	KRT78	16	51519012	51529045	0.00395
12	KRT8	14	51577237	51585127	0.00397
5	C5orf39	5	43074938	43076098	0.00398
1	PLA2G4A	20	185064654	185224736	0.00412
11	OR8D2	10	123694367	123695303	0.00423
22	LIF	7	28966442	28972748	0.00424

References

1. de Jong,P.T. Age-related macular degeneration. *N. Engl. J. Med.* **355**, 1474-1485 (2006).

2. van Lookeren,C.M., LeCouter,J., Yaspan,B.L., & Ye,W. Mechanisms of age-related macular degeneration and therapeutic opportunities. *J. Pathol.* **232**, 151-164 (2014).

3. Hageman,G.S. *et al.* 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**, 705-732 (2001).

4. Abdelsalam,A., Del,P.L., & Zarbin,M.A. Drusen in age-related macular degeneration: pathogenesis, natural course, and laser photocoagulation-induced regression. *Surv. Ophthalmol.* **44**, 1-29 (1999).

5. Hageman,G.S. *et al.* 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**, 705-732 (2001).

6. Crabb,J.W. *et al.* Drusen proteome analysis: an approach to the etiology of age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A* **99**, 14682-14687 (2002).

7. Klein,R. *et al.* Markers of inflammation, oxidative stress, and endothelial dysfunction and the 20-year cumulative incidence of early age-related macular degeneration: the Beaver Dam Eye Study. *JAMA Ophthalmol.* **132**, 446-455 (2014).

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

9. Sarks,J.P., Sarks,S.H., & Killingsworth,M.C. Evolution of geographic atrophy of the retinal pigment epithelium. *Eye (Lond)* **2 (Pt 5)**, 552-577 (1988).

10. Klein,R., Wang,Q., Klein,B.E., Moss,S.E., & Meuer,S.M. The relationship of age-related maculopathy, cataract, and glaucoma to visual acuity. *Invest Ophthalmol. Vis. Sci.* **36**, 182-191 (1995).

11. Holz,F.G. *et al.* Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am. J. Ophthalmol.* **143**, 463-472 (2007).

12. Sunness,J.S. *et al.* The long-term natural history of geographic atrophy from age-related macular degeneration: enlargement of atrophy and implications for interventional clinical trials. *Ophthalmology* **114**, 271-277 (2007).

13. Ambati,J. & Fowler,B.J. Mechanisms of age-related macular degeneration 12. *Neuron* **75**, 26-39 (2012).

14. Ambati,J., Ambati,B.K., Yoo,S.H., Ianchulev,S., & Adamis,A.P. Age-related macular degeneration: etiology, pathogenesis, and therapeutic strategies. *Surv. Ophthalmol.* **48**, 257-293 (2003).

15. Husain,D., Ambati,B., Adamis,A.P., & Miller,J.W. Mechanisms of age-related macular degeneration. *Ophthalmol. Clin. North Am.* **15**, 87-91 (2002).

16. Klein,R. *et al.* Prevalence of age-related macular degeneration in 4 racial/ethnic groups in the multi-ethnic study of atherosclerosis. *Ophthalmology* **113**, 373-380 (2006).

17. Klein,R. *et al.* Age-related maculopathy in a multiracial United States population: the National Health and Nutrition Examination Survey III. *Ophthalmology* **106**, 1056-1065 (1999).

18. Klein,R., Klein,B.E., & Cruickshanks,K.J. The prevalence of age-related maculopathy by geographic region and ethnicity. *Prog. Retin. Eye Res.* **18**, 371-389 (1999).

19. Vingerling,J.R., Klaver,C.C., Hofman,A., & de Jong,P.T. Epidemiology of age-related maculopathy. *Epidemiol. Rev.* **17**, 347-360 (1995).

20. Smith,W. *et al.* Risk factors for age-related macular degeneration: Pooled findings from three continents. *Ophthalmology* **108**, 697-704 (2001).

21. Wang,J.J. *et al.* Ten-year incidence and progression of age-related maculopathy: the blue Mountains Eye Study. *Ophthalmology* **114**, 92-98 (2007).

22. Klein,R. *et al.* The prevalence of age-related macular degeneration and associated risk factors. *Arch. Ophthalmol.* **128**, 750-758 (2010).

23. Friedman,D.S. *et al.* Prevalence of age-related macular degeneration in the United States. *Arch. Ophthalmol.* **122**, 564-572 (2004).

24. Smith,W., Mitchell,P., & Wang,J.J. Gender, oestrogen, hormone replacement and age-related macular degeneration: results from the Blue Mountains Eye Study. *Aust. N. Z. J. Ophthalmol.* **25 Suppl 1**, S13-S15 (1997).

25. Klein,R., Klein,B.E., & Moss,S.E. Relation of smoking to the incidence of age-related maculopathy. The Beaver Dam Eye Study. *Am. J. Epidemiol.* **147**, 103-110 (1998).

26. Mitchell,P., Wang,J.J., Smith,W., & Leeder,S.R. Smoking and the 5-year incidence of age-related maculopathy: the Blue Mountains Eye Study. *Arch. Ophthalmol.* **120**, 1357-1363 (2002).

27. Klein,R., Knudtson,M.D., Cruickshanks,K.J., & Klein,B.E. Further observations on the association between smoking and the long-term incidence and progression of age-related macular degeneration: the Beaver Dam Eye Study. *Arch. Ophthalmol.* **126**, 115-121 (2008).

28. Smith,W., Mitchell,P., & Leeder,S.R. Smoking and age-related maculopathy. The Blue Mountains Eye Study. *Arch. Ophthalmol.* **114**, 1518-1523 (1996).

29. Klein,R., Klein,B.E., Linton,K.L., & DeMets,D.L. The Beaver Dam Eye Study: the relation of age-related maculopathy to smoking. *Am. J. Epidemiol.* **137**, 190-200 (1993).

30. Khan,J.C. *et al.* Smoking and age related macular degeneration: the number of pack years of cigarette smoking is a major determinant of risk for both geographic atrophy and choroidal neovascularisation. *Br. J. Ophthalmol.* **90**, 75-80 (2006).

31. Tan,J.S., Mitchell,P., Smith,W., & Wang,J.J. Cardiovascular risk factors and the long-term incidence of age-related macular degeneration: the Blue Mountains Eye Study. *Ophthalmology* **114**, 1143-1150 (2007).

32. Smith,W. *et al.* Risk factors for age-related macular degeneration: Pooled findings from three continents. *Ophthalmology* **108**, 697-704 (2001).

33. Chew,E.Y. *et al.* Risk of advanced age-related macular degeneration after cataract surgery in the Age-Related Eye Disease Study: AREDS report 25. *Ophthalmology* **116**, 297-303 (2009).

34. Tomany,S.C., Cruickshanks,K.J., Klein,R., Klein,B.E., & Knudtson,M.D. Sunlight and the 10-year incidence of age-related maculopathy: the Beaver Dam Eye Study. *Arch. Ophthalmol.* **122**, 750-757 (2004).

35. Pham,T.Q., Rochtchina,E., Mitchell,P., Smith,W., & Wang,J.J. Sunlight-related factors and the 10-year incidence of age-related maculopathy. *Ophthalmic Epidemiol.* **16**, 136-141 (2009).

36. Cruickshanks,K.J., Klein,R., Klein,B.E., & Nondahl,D.M. Sunlight and the 5-year incidence of early age-related maculopathy: the beaver dam eye study. *Arch. Ophthalmol.* **119**, 246-250 (2001).

37. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Arch. Ophthalmol.* **119**, 1417-1436 (2001).

38. Lutein + zeaxanthin and omega-3 fatty acids for age-related macular degeneration: the Age-Related Eye Disease Study 2 (AREDS2) randomized clinical trial. *JAMA* **309**, 2005-2015 (2013).

39. Rosenfeld,P.J. *et al.* Ranibizumab for neovascular age-related macular degeneration. *N. Engl. J. Med.* **355**, 1419-1431 (2006).

40. Seddon,J.M., Ajani,U.A., & Mitchell,B.D. Familial aggregation of age-related maculopathy. *Am. J. Ophthalmol.* **123**, 199-206 (1997).

41. Martin,D.F. *et al.* Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N. Engl. J. Med.* **364**, 1897-1908 (2011).

42. Seddon,J.M., Cote,J., Page,W.F., Aggen,S.H., & Neale,M.C. The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences. *Arch. Ophthalmol.* **123**, 321-327 (2005).

43. Hammond,C.J. *et al.* Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology* **109**, 730-736 (2002).

44. Klein,M.L. *et al.* Age-related macular degeneration. Clinical features in a large family and linkage to chromosome 1q. *Arch. Ophthalmol.* **116**, 1082-1088 (1998).

45. Kenealy,S.J. *et al.* Linkage analysis for age-related macular degeneration supports a gene on chromosome 10q26. *Mol. Vis.* **10**, 57-61 (2004).

46. Majewski,J. *et al.* Age-related macular degeneration--a genome scan in extended families. *Am. J. Hum. Genet.* **73**, 540-550 (2003).

47. Seddon,J.M., Santangelo,S.L., Book,K., Chong,S., & Cote,J. A genomewide scan for age-related macular degeneration provides evidence for linkage to several chromosomal regions. *Am. J. Hum. Genet.* **73**, 780-790 (2003).

48. Weeks,D.E. *et al.* A full genome scan for age-related maculopathy. *Hum. Mol. Genet.* **9**, 1329-1349 (2000).

49. Weeks,D.E. *et al.* Age-related maculopathy: an expanded genome-wide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. *Am. J. Ophthalmol.* **132**, 682-692 (2001).

50. Jakobsdottir,J. *et al.* Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am. J. Hum. Genet.* **77**, 389-407 (2005).

51. A haplotype map of the human genome. *Nature* **437**, 1299-1320 (2005).

52. The International HapMap Project. *Nature* **426**, 789-796 (2003).

53. Edwards,A.O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421-424 (2005).

54. Hageman,G.S. *et al.* 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**, 7227-7232 (2005).

55. Haines,J.L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419-421 (2005).

56. Klein,R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389 (2005).

57. Rivera,A. *et al.* 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**, 3227-3236 (2005).

58. Li,M. *et al. CFH* haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat. Genet.* **38**, 1049-1054 (2006).

59. Clark,S.J. *et al.* Identification of factor H-like protein 1 as the predominant complement regulator in Bruch's membrane: implications for age-related macular degeneration. *J. Immunol.* **193**, 4962-4970 (2014).

60. Clark,S.J. *et al.* Impaired binding of the age-related macular degeneration-associated complement factor H 402H allotype to Bruch's membrane in human retina. *J. Biol. Chem.* **285**, 30192-30202 (2010).

61. Yates,J.R. *et al.* Complement C3 variant and the risk of age-related macular degeneration. *N. Engl. J. Med.* **357**, 553-561 (2007).

62. Maller,J. *et al.* Common variation in three genes, including a noncoding variant in *CFH*, strongly influences risk of age-related macular degeneration. *Nat. Genet.* **38**, 1055-1059 (2006).

63. Gold,B. *et al.* Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat. Genet.* **38**, 458-462 (2006).

64. Spencer,K.L. *et al.* Deletion of *CFH*R3 and *CFH*R1 genes in age-related macular degeneration. *Hum. Mol. Genet.* **17**, 971-977 (2008).

65. Fritsche,L.G. *et al.* An imbalance of human complement regulatory proteins *CFH*R1, *CFH*R3 and factor H influences risk for age-related macular degeneration (AMD)

3. *Hum. Mol. Genet.* **19**, 4694-4704 (2010).

66. Wang,G. *et al.* Localization of age-related macular degeneration-associated ARMS2 in cytosol, not mitochondria. *Invest Ophthalmol. Vis. Sci.* **50**, 3084-3090 (2009).

67. Francis,P.J., Zhang,H., Dewan,A., Hoh,J., & Klein,M.L. Joint effects of polymorphisms in the HTRA1, LOC387715/ARMS2, and *CFH* genes on AMD in a Caucasian population. *Mol. Vis.* **14**, 1395-1400 (2008).

68. Yang,Z. *et al.* A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science* **314**, 992-993 (2006).

69. Fritsche,L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433-2 (2013).

70. Chen,Y. *et al.* Assessing susceptibility to age-related macular degeneration with genetic markers and environmental factors. *Arch. Ophthalmol.* **129**, 344-351 (2011).

71. Chowers,I. *et al.* Sequence variants in HTRA1 and LOC387715/ARMS2 and phenotype and response to photodynamic therapy in neovascular age-related macular degeneration in populations from Israel. *Mol. Vis.* **14**, 2263-2271 (2008).

72. Kanda,A. *et al.* Age-related macular degeneration-associated variants at chromosome 10q26 do not significantly alter ARMS2 and HTRA1 transcript levels in the human retina. *Mol. Vis.* **16**, 1317-1323 (2010).

73. Yang,Z. *et al.* A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science* **314**, 992-993 (2006).

74. Yang,J. *et al.* Validation of genome-wide association study (GWAS)-identified disease risk alleles with patient-specific stem cell lines. *Hum. Mol. Genet.* **23**, 3445-3455 (2014).

75. Tuo,J. *et al.* The HtrA1 promoter polymorphism, smoking, and age-related macular degeneration in multiple case-control samples. *Ophthalmology* **115**, 1891-1898 (2008).

76. Fritsche,L.G. *et al.* Age-related macular degeneration is associated with an unstable ARMS2 (LOC387715) mRNA. *Nat. Genet.* **40**, 892-896 (2008).

77. Wang,G. *et al.* Analysis of the indel at the ARMS2 3'UTR in age-related macular degeneration. *Hum. Genet.* **127**, 595-602 (2010).

78. Cirulli,E.T. & Goldstein,D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415-425 (2010).

79. Manolio,T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).

80. van de Ven,J.P. *et al.* A functional variant in the CFI gene confers a high risk of age-related macular degeneration. *Nat. Genet.* **45**, 813-817 (2013).

81. Helgason,H. *et al.* A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. *Nat. Genet.* **45**, 1371-1374 (2013).

82. Seddon,J.M. *et al.* Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat. Genet.* **45**, 1366-1370 (2013).

83. Beachy,A., Hershberger,E., Davidhizar,R., & Giger,J.N. Cultural implications for nursing care of the Amish. *J. Cult. Divers.* **4**, 118-126 (1997)

84. Amish Heritage Committee *Amish and Mennonites in Eastern Elkhart & LaGrange Counties, Indiana 1841-1991*(Amish Heritage Committee, Goshen, 2009).

85. Agarwala,R., Biesecker,L.G., Tomlin,J.F., & Schaffer,A.A. Towards a complete North American Anabaptist genealogy: A systematic approach to merging partially overlapping genealogy resources. *Am. J. Med. Genet.* **86**, 156-161 (1999).

86. Agarwala,R., Biesecker,L.G., & Schaffer,A.A. Anabaptist genealogy database. *Am. J. Med. Genet. C. Semin. Med. Genet.* **121C**, 32-37 (2003).

87. Hahs,D.W. *et al.* A genome-wide linkage analysis of dementia in the Amish. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **141B**, 160-166 (2006).

88. McCauley,J.L. *et al.* Combinatorial Mismatch Scan (CMS) for loci associated with dementia in the Amish. *BMC. Med. Genet.* **7**, 19 (2006).

89. Cummings,A.C. *et al.* Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene. *Ann. Hum. Genet.* **76**, 342-351 (2012).

90. Bartlett,H. & Eperjesi,F. Use of fundus imaging in quantification of age-related macular change. *Surv. Ophthalmol.* **52**, 655-671 (2007).

91. Klein,R. *et al.* The Wisconsin age-related maculopathy grading system. *Ophthalmology* **98**, 1128-1134 (1991).

92. The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1. *Control Clin. Trials* **20**, 573-600 (1999).

93. Spencer,K.L. *et al.* Haplotypes spanning the complement factor H gene are protective against age-related macular degeneration. *Invest Ophthalmol. Vis. Sci.* **48**, 4277-4283 (2007).

94. van Grinsven,M.J. *et al.* Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images. *Invest Ophthalmol. Vis. Sci.* **54**, 3019-3027 (2013).

95. van Grinsven,M.J. *et al.* Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images. *Invest Ophthalmol. Vis. Sci.* **54**, 3019-3027 (2013).

96. Clemons,T.E., Milton,R.C., Klein,R., Seddon,J.M., & Ferris,F.L., III 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**, 533-539 (2005).

97. D.Haung *Retinal Imaging*(Elsevier,2006).

98. Klein,R. *et al.* Fifteen-year cumulative incidence of age-related macular degeneration: the Beaver Dam Eye Study. *Ophthalmology* **114**, 253-262 (2007).

99. Evans,J.R., Fletcher,A.E., & Wormald,R.P. 28,000 Cases of age related macular degeneration causing visual loss in people aged 75 years and above in the United Kingdom may be attributable to smoking. *Br. J. Ophthalmol.* **89**, 550-553 (2005).

100. Khan,J.C. *et al.* Smoking and age related macular degeneration: the number of pack years of cigarette smoking is a major determinant of risk for both geographic atrophy and choroidal neovascularisation. *Br. J. Ophthalmol.* **90**, 75-80 (2006).

101. Klein,R., Klein,B.E., Tomany,S.C., Meuer,S.M., & Huang,G.H. Ten-year incidence and progression of age-related maculopathy: The Beaver Dam eye study. *Ophthalmology* **109**, 1767-1779 (2002).

102. Klein,R., Knudtson,M.D., Lee,K.E., Gangnon,R.E., & Klein,B.E. Age-period-cohort effect on the incidence of age-related macular degeneration: the Beaver Dam Eye Study. *Ophthalmology* **115**, 1460-1467 (2008).

103. Schaumberg,D.A., Christen,W.G., Hankinson,S.E., & Glynn,R.J. Body mass index and the incidence of visually significant age-related maculopathy in men. *Arch. Ophthalmol.* **119**, 1259-1265 (2001).

104. Thornton,J. *et al.* Smoking and age-related macular degeneration: a review of association. *Eye (Lond)* **19**, 935-944 (2005).

105. Thylefors,B., Negrel,A.D., Pararajasegaram,R., & Dadzie,K.Y. Global data on blindness. *Bull. World Health Organ* **73**, 115-121 (1995).

106. Maller,J.B. *et al.* Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat. Genet.* **39**, 1200-1201 (2007).

107. Schmidt,S. *et al.* Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am. J. Hum. Genet.* **78**, 852-864 (2006).

108. Souied,E.H. *et al.* Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. *Mol. Vis.* **11**, 1135-1140 (2005).

109. Spencer,K.L. *et al.* Protective effect of complement factor B and complement component 2 variants in age-related macular degeneration. *Hum. Mol. Genet.* **16**, 1986-1992 (2007).

110. Spencer,K.L. *et al.* C3 R102G polymorphism increases risk of age-related macular degeneration. *Hum. Mol. Genet.* **17**, 1821-1824 (2008).

111. Clemons,T.E., Milton,R.C., Klein,R., Seddon,J.M., & Ferris,F.L., III 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**, 533-539 (2005).

112. Dreyhaupt,J. *et al.* Modelling the natural history of geographic atrophy in patients with age-related macular degeneration. *Ophthalmic Epidemiol.* **12**, 353-362 (2005).

113. Farwick,A., Wellmann,J., Stoll,M., Pauleikhoff,D., & Hense,H.W. Susceptibility genes and progression in age-related maculopathy: a study of single eyes. *Invest Ophthalmol. Vis. Sci.* **51**, 731-736 (2010).

114. Fleckenstein,M. *et al.* Concordance of disease progression in bilateral geographic atrophy due to AMD. *Invest Ophthalmol. Vis. Sci.* **51**, 637-642 (2010).

115. Fleckenstein,M. *et al.* Tracking progression with spectral-domain optical coherence tomography in geographic atrophy caused by age-related macular degeneration. *Invest Ophthalmol. Vis. Sci.* **51**, 3846-3852 (2010).

116. Fleckenstein,M. *et al.* Progression of age-related geographic atrophy: role of the fellow eye. *Invest Ophthalmol. Vis. Sci.* **52**, 6552-6557 (2011).

117. Francis,P.J., Hamon,S.C., Ott,J., Weleber,R.G., & Klein,M.L. Polymorphisms in C2, CFB and C3 are associated with progression to advanced age related macular degeneration associated with visual loss. *J. Med. Genet.* **46**, 300-307 (2009).

118. Oliver-Fernandez,A. *et al.* Progression of visual loss and time between initial assessment and treatment of wet age-related macular degeneration. *Can. J. Ophthalmol.* **40**, 313-319 (2005).

119. Seddon,J.M., Cote,J., Davis,N., & Rosner,B. Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio. *Arch. Ophthalmol.* **121**, 785-792 (2003).

120. Wang,J.J. *et al.* Combined effects of complement factor H genotypes, fish consumption, and inflammatory markers on long-term risk for age-related macular degeneration in a cohort. *Am. J. Epidemiol.* **169**, 633-641 (2009).

121. Seddon,J.M., Reynolds,R., Yu,Y., Daly,M.J., & Rosner,B. Risk models for progression to advanced age-related macular degeneration using demographic, environmental, genetic, and ocular factors. *Ophthalmology* **118**, 2203-2211 (2011).

122. Schmidt,S. *et al.* Association of the apolipoprotein E gene with age-related macular degeneration: possible effect modification by family history, age, and gender. *Mol. Vis.* **6**, 287-293 (2000).

123. Seddon,J.M., Sharma,S., & Adelman,R.A. Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology* **113**, 260-266 (2006).

124. Fritsche,L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433-2 (2013).

125. Hoffman,J.D. *et al.* Rare complement factor H variant associated with age-related macular degeneration in the Amish. *Invest Ophthalmol. Vis. Sci.* **55**, 4455-4460 (2014).

126. Purcell,S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).

127. Yaspan,B.L. *et al.* Genetic analysis of biological pathway data through genomic randomization. *Hum. Genet.* **129**, 563-571 (2011).

128. Diniz,B. *et al.* Drusen and RPE atrophy automated quantification by optical coherence tomography in an elderly population. *Eye (Lond)*(2014).

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

130. Klein,R. *et al.* Markers of inflammation, oxidative stress, and endothelial dysfunction and the 20-year cumulative incidence of early age-related macular degeneration: the Beaver Dam Eye Study. *JAMA Ophthalmol.* **132**, 446-455 (2014).

131. Klein,R. *et al.* Prevalence of age-related macular degeneration in the US population. *Arch. Ophthalmol.* **129**, 75-80 (2011).

132. Haung,D. *Retinal Imaging*(Elsevier,2006).

133. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) date(02,2013) accessed.

Ref Type: Generic

134. Kenealy,S.J. *et al.* Linkage analysis for age-related macular degeneration supports a gene on chromosome 10q26. *Mol. Vis.* **10**, 57-61 (2004).

135. Fritsche,L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433-2 (2013).

136. Manolio,T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).

137. Zhan,X. *et al.* Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat. Genet.* **45**, 1375-1379 (2013).

138. Raychaudhuri,S. *et al.* A rare penetrant mutation in *CFH* confers high risk of age-related macular degeneration. *Nat. Genet.* **43**, 1232-1236 (2011).

139. Hostetler,J. *Amish Society*(Johns Hopkins University Press, Baltimore, MD, 1993).

140. Agarwala,R., Biesecker,L.G., & Schaffer,A.A. Anabaptist genealogy database. *Am. J. Med. Genet. C. Semin. Med. Genet.* **121C**, 32-37 (2003).

141. Agarwala,R., Schaffer,A.A., & Tomlin,J.F. Towards a complete North American Anabaptist Genealogy II: analysis of inbreeding. *Hum. Biol.* **73**, 533-545 (2001).

142. McCauley,J.L. *et al.* Combinatorial Mismatch Scan (CMS) for loci associated with dementia in the Amish. *BMC. Med. Genet.* **7**, 19 (2006).

143. Lee,W.J., Pollin,T.I., O'Connell,J.R., Agarwala,R., & Schaffer,A.A. PedHunter 2.0 and its usage to characterize the founder structure of the Old Order Amish of Lancaster County. *BMC. Med. Genet.* **11**, 68 (2010).

144. Li,H. & Durbin,R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754-1760 (2009).

145. DePristo,M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498 (2011).

146. Thornton,T. & McPeek,M.S. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am. J. Hum. Genet.* **81**, 321-337 (2007).

147. Mitchell,P., Smith,W., Attebo,K., & Wang,J.J. Prevalence of age-related maculopathy in Australia. The Blue Mountains Eye Study. *Ophthalmology* **102**, 1450-1460 (1995).

148. Liu,F., Kirichenko,A., Axenovich,T.I., van Duijn,C.M., & Aulchenko,Y.S. An approach for cutting large and complex pedigrees for linkage analysis. *Eur. J. Hum. Genet.* **16**, 854-860 (2008).

149. Abecasis,G.R., Cherny,S.S., Cookson,W.O., & Cardon,L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97-101 (2002).

150. Cummings,A.C. *et al.* Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene. *Ann. Hum. Genet.* **76**, 342-351 (2012).

151. Pangburn,M.K., Pangburn,K.L., Koistinen,V., Meri,S., & Sharma,A.K. Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement. *J. Immunol.* **164**, 4742-4751 (2000).

152. Sharma,A.K. & Pangburn,M.K. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl. Acad. Sci. U. S. A* **93**, 10996-11001 (1996).

153. Chen,Y. *et al.* Assessing susceptibility to age-related macular degeneration with genetic markers and environmental factors. *Arch. Ophthalmol.* **129**, 344-351 (2011).

154. Hubbard,L.D. *et al.* Brightness, contrast, and color balance of digital versus film retinal images in the age-related eye disease study 2. *Invest Ophthalmol. Vis. Sci.* **49**, 3269-3282 (2008).

155. Weeks,D.E. *et al.* Age-related maculopathy: an expanded genome-wide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. *Am. J. Ophthalmol.* **132**, 682-692 (2001).

156. Lozano,E., Joller,N., Cao,Y., Kuchroo,V.K., & Hafler,D.A. The CD226/CD155 interaction regulates the proinflammatory (Th1/Th17)/anti-inflammatory (Th2) balance in humans. *J. Immunol.* **191**, 3673-3680 (2013).

157. Yehoshua,Z. *et al.* Comparison of drusen area detected by spectral domain optical coherence tomography and color fundus imaging. *Invest Ophthalmol. Vis. Sci.* **54**, 2429-2434 (2013).

158. Knudtson,M.D. *et al.* Location of lesions associated with age-related maculopathy over a 10-year period: the Beaver Dam Eye Study. *Invest Ophthalmol. Vis. Sci.* **45**, 2135-2142 (2004).