Am. J. Hum. Genet. 73:540–550, 2003

Age-Related Macular Degeneration—a Genome Scan in Extended Families

Jacek Majewski,[1] Dennis W. Schultz,[2] Richard G. Weleber,[2,3] Mitchell B. Schain,[2] Albert O. Edwards,[5] Tara C. Matise,[1,*] Ted S. Acott,[2,4] Jurg Ott,[1] and Michael L. Klein,[2]

[1]Laboratory of Statistical Genetics, Rockefeller University, New York; [2]Macular Degeneration Center, Casey Eye Institute, Department of Ophthalmology, Oregon Health and Science University, and Departments of [3]Molecular and Medical Genetics and [4]Biochemistry and Molecular Biology, Oregon Health and Science University, Portland; and [5]Department of Ophthalmology, University of Texas Southwestern Medical Center, Dallas

We performed a genomewide scan and genetic linkage analysis, to identify loci associated with age-related macular degeneration (AMD). We collected 70 families, ranging from small nuclear families to extended multigenerational pedigrees and consisting of a total of 344 affected and 217 unaffected members available for genotyping. We performed linkage analyses using parametric and allele-sharing models. We performed the analyses on the complete pedigrees but also subdivided the families into nuclear pedigrees. Finally, to dissect potential genetic factors responsible for differences in disease manifestation, we stratified the sample by two major AMD phenotypes (neovascular AMD and geographic atrophy) and by age of affected family members at the time of our evaluation. We have previously demonstrated linkage between AMD and 1q25-31 in a single large family. In the combined sample, we have detected the following loci with scores exceeding a LOD = 2 cutoff under at least one of the models considered: 1q31 (HLOD = 2.07 at D1S518), 3p13 (HLOD = 2.19 at D3S1304/D3S4545), 4q32 (HLOD = 2.66 at D4S2368, for the subset of families with predominantly dry AMD), 9q33 ($LOD_{Z_{lr}}$ = 2.01 at D9S930/D9S934), and 10q26 (HLOD = 3.06 at D10S1230). Using correlation analysis, we have found a statistically significant correlation between LOD scores at 3p13 and 10q26, providing evidence for epistatic interactions between the loci and, hence, a complex basis of AMD. Our study has identified new loci that should be considered in future mapping and mutational analyses of AMD and has strengthened the evidence in support of loci suggested by other studies.

Introduction

Age-related macular degeneration (AMD [MIM 603075]) is a leading cause of blindness in the United States and other Western countries (Klein et al. 1992; Evans and Wormald 1996). In recent years, evidence for a significant genetic component has prompted a search for genes that may cause or increase susceptibility to AMD. Previous studies have consisted primarily of case-control methods (Allikmets et al. 1997) or linkage studies employing small families (Weeks et al. 2001).

In the present article, we describe our findings from a genomewide scan of extended families with AMD. A large part of our data set consists of multigenerational families. Families of this type have previously been successfully employed in the identification of at least one candidate AMD locus, 1q25-31 (Klein et al. 1998), which was later confirmed by an independent study (Weeks et al. 2001). In contrast to other studies, which have focused on affected sibs and relatively small nuclear families (Weeks et al. 2000, 2001; Haines et al. 2002; Iyengar et al. 2002), some of the families that we have studied are large enough to individually attain, under a dominant genetic model, maximum LOD scores exceeding the generally accepted cutoff of 3. Previous AMD studies (Weeks et al. 2000, 2001; Haines et al. 2002; Iyengar et al. 2002) have met with limited success, possibly because of a complex genetic basis of the disease or because of a high degree of genetic heterogeneity (i.e., a large number of independent loci and mutations being responsible for the observed phenotype). Because of the large size of the families that we studied, if the genetic basis of the disease is relatively close to dominant inheritance and only one locus is segregating in each family, then we should be able to detect linkage within individual families even in the presence of genetic heterogeneity. Our combined sample should also be large enough to identify loci that contribute to possible complex inheritance patterns of AMD. In the present article, we identify several candidate loci, some of which constitute novel findings and others of which have been implicated in previous studies. In view

Received March 21, 2003; accepted for publication June 23, 2003; electronically published July 25, 2003.

Address for correspondence and reprints: Dr. Michael L. Klein, Macular Degeneration Center, Casey Eye Institute, Oregon Health and Science University, 3375 SW Terwilliger Boulevard, Portland, OR 97201-4197. E-mail: kleinm@ohsu.edu

* Present affiliation: Department of Genetics, Rutgers University, Piscataway, NJ.

0002-9297/2003/7303-0008$15.00

of our results, we also discuss the possible etiology of AMD.

Families and Methods

Families

The study population consisted of 561 members of 70 families recruited in the Pacific Northwest region of the United States. There were 344 affected and 217 unaffected individuals. Each family had 3–14 (average 4.9) living family members with AMD. Families were recruited initially from M.L.K.'s practice and later via advertisements in various newspapers in the Northwest and organizations serving the blind. Extensive recruitment efforts to recruit new families included announcements and presentations at local meetings, letters to eye-care providers, articles in local senior-citizen newspapers, articles in institutional news publications, and paid recruitment advertisements in major newspapers throughout the Pacific Northwest. We also had good success in the identification of new families after we were given permission to contact members of the Oregon Commission for the Blind. All families were of northern European descent. Pedigrees had to have a minimum of three living members who met our definition of AMD to be considered as a family. AMD was defined as follows: (1) geographic atrophy or choroidal neovascularization accompanied by drusen in the macula ("late age-related maculopathy" [ARM]) or (2) extensive large drusen >393,744 μm^2 in area (0.2 disc areas). In the Beaver Dam Eye Study, this amount of drusen represented the highest-risk category for the development of late ARM (55.3% in 10 years) (Klein et al. 2002). Unaffected individuals (Klein et al. 1998) did not have these features, and those with no drusen >125 μm in minimum diameter were considered as definitely unaffected. We documented affected status of family members with retinal photography.

Genotyping

Genotyping was performed in two successive stages. The initial set of 21 families with AMD, comprising only affected individuals, was genotyped by the National Heart, Lung, and Blood Institute (NHLBI) Mammalian Genotyping Service, using a full-genome set of 364 markers from Marshfield screening set 9. While the first analysis was under way, additional families with AMD, along with additional individuals from the families already under investigation, were ascertained. The total sample size was thus increased to 70 informative pedigrees, consisting of a total of 344 affected and 217 unaffected members available for genotyping.

The second set of families, which included unaffected individuals, was again genotyped by the Marshfield Foundation, but the NHLBI Mammalian Genotyping Service this time used a newer panel of markers, screening set 10. The data from the two genome scans were subsequently combined and used in linkage analysis.

Linkage Analysis

Initially, we analyzed the original set of 21 families. We checked the genotyping data for errors by using the PedManager (available from M. P. Reeve of the Massachusetts Institute of Technology [mpreeve@genome.wi.mit.edu]) and PedCheck (O'Connell and Weeks 1998) programs. In many cases, the error could be eliminated by untyping a single individual in predominantly multiple-sib families. In the remaining cases, the entire nuclear family in which an inconsistency was found was untyped. The Marshfield genotyping software indicated several mispaternity cases, which we then handled by including an additional father in the pedigree files. We analyzed the data by using the Analyze/Linkage software package (Cottingham et al. 1993; Terwilliger and Ott 1994), for two-point analysis, and the Allegro software (Gudbjartsson et al. 2000), for multipoint and nonparametric analyses. Later, when a number of additional families had been collected, they were added to the sample, yielding a total of 70 families analyzed.

The combined data included results from two separate genotyping efforts. Several markers differed between the two genotyping panels (screening sets 9 and 10); some were present only in the first scan, and others were present only in the second scan. In the second scan, many of the original markers were replaced by more polymorphic, or more consistently amplified, markers at the same genetic position. In such cases, we obtained the physical order of the markers from the human genomic sequence assembly at the University of California at Santa Cruz (Kent et al. 2002) and assumed an arbitrary genetic separation of 0.01 cM for multipoint linkage analysis. In addition, within families, some individuals were often typed using screening set 9, and others were typed using screening set 10. In effect, to extract full information from the pedigrees, we relied on multipoint, rather than two-point, analyses. Finally, even for markers shared between the two scans, there remained four obvious inconsistencies in allele sizing. We were able to correct those discrepancies by recalibrating allele sizes to fit consistent Mendelian transmission.

We performed linkage analysis under three models (for summary, see table 1). Model 1 is a parametric model with a dominant mode of inheritance, 0.01 frequency of the disease allele, and a set of age-dependent penetrances determined by the frequency of AMD in various age groups within families, as described by Klein et al. (1998). For each age group, the phenocopy rate is 10% of its associated penetrance. We used the following phenocopy and penetrance values, with the three val-

Table 1

Analysis Methods

Model	Diagnostic Scheme		
	A. All Affected and Unaffected Individuals	B. All Affected and Definitely Unaffected Individuals	C. All Affected and Definitely Unaffected Individuals (Nuclear Families)
1. Parametric with age-dependent penetrances	1A	1B	…
2. Parametric with constant high penetrance and low phenocopy rate	2A	2B	2C
3. Nonparametric allele sharing (Z_{lr})	3A	…	3C

ues given in order corresponding to noncarriers of the disease allele, heterozygous carriers, and homozygous carriers: (0.0001, 0.001, 0.001), (0.001, 0.01, 0.01), (0.009, 0.09, 0.09), (0.042, 0.42, 0.42), and (0.095, 0.95, 0.95), for the respective age groups of <50, 50–54, 55–64, 65–74, and ⩾75 years.

Although the model described in the preceding paragraph (model 1) is derived in a rigorous manner, in general, the use of genetic models that include a large percentage of phenocopies (affected noncarriers) results in low power to detect linkage. Hence, we also used model 2, a parametric approach that assumes a relatively low phenocopy rate of 0.05, as well as relatively high penetrances of 0.95 for both hetero- and homozygous individuals. Although the parameters of the model are arbitrary, faced with an uncertain mode of inheritance of AMD, we used this model in order to increase the statistical power of the analysis. For both parametric models, we assumed a 0.01 frequency of the disease allele; this relatively high frequency is justified in view of the high population prevalence of AMD. Note that, although the use of a misspecified genetic model does not lead to an increase in false-positive rates (Williamson and Amos 1995), the use of multiple models will result in an overall decreased significance of each analysis.

In view of the uncertain mode of inheritance of the disease, we also used model 3, a nonparametric approach based on an allele-sharing statistic, the exponential pairs Z_{lr} score. Among allele-sharing approaches, the above statistic has been found to perform well for a variety of underlying disease models (McPeek 1999).

Care should be taken when comparing the results of the parametric and nonparametric analyses. The parametric LOD scores allowing for genetic heterogeneity (HLOD scores), which are reported here, require the estimation of the heterogeneity parameter (proportion of families segregating a given locus) and have >1 df. Ott (1999) recommends increasing the significance cutoff required for LOD scores by 0.3, to account for the additional parameter—that is, HLOD = 3.3 corresponds to the statistical significance of LOD = 3. Thus, pointwise results of parametric (HLOD) and nonparametric ($LOD_{Z_{lr}}$) analyses are not equivalent. However, in their classic article concerning genomewide significance levels, Lander and Kruglyak (1995) recommend a genomewide 3.3 cutoff for parametric LOD (and, therefore, a 3.6 threshold for HLOD) and a 3.6 cutoff for allele-sharing methods based on sib pairs. The latter threshold is even higher for allele sharing between other relative pairs. Thus, when considered in a genomewide context, the two statistics (HLOD and $LOD_{Z_{lr}}$) may actually be comparable.

We performed the above analyses under two different diagnostic schemes. Scheme A included the entire collection of affected, unaffected, and unclassified individuals. Scheme B included affected and unclassified individuals and only those unaffected individuals classified as definitely unaffected (all other unaffected individuals were included in the analysis but with an unknown affection status).

It has been suggested elsewhere (Badner et al. 1998) that, for the analysis of complex traits caused by common alleles, extended pedigrees may not be the optimal units for linkage analysis and that, in some cases, the power obtained from nuclear families may be greater than from large pedigrees. Hence, we also used scheme C, which followed the diagnostic criteria of scheme B (above) but also involved subdividing the pedigrees into nuclear families (72 informative nuclear pedigrees, with averages of 3.1 affected and 0.5 unaffected sibs). This approach should be more effective in the presence of several disease loci segregating in large pedigrees and is more similar to approaches used by other groups (Weeks et al. 2000, 2001; Iyengar et al. 2002), in which collections of sib pairs or much smaller nuclear families are often analyzed.

In addition, we considered the possibility that AMD subgroups may exist that have distinct genetic influences. We therefore performed additional analyses on samples stratified according to two important criteria: (1) advanced AMD phenotype (i.e., dry [geographic atrophy] vs. wet [neovascular] AMD) and (2) age at ascertainment (i.e., early vs. late). Thus, the first stratification involved subdividing our sample into families in which members had advanced AMD that was predominantly dry (>50% of the affected individuals, with a total of 27 families) or predominantly wet (43 families). The second stratification included families com-

Table 2

Multipoint LOD Scores Exceeding the 1.5 Cutoff in Individual Families (Model 2B)

Marker(s)	Position (cM)	Family	LOD	*P*	Reference(s)
Chromosome 1:					
D1S3723/D1S534	150.13	AMD199	1.58	.24	
D1S1589	192.05	AMD159	2.59	<.01	Weeks et al. 2001
D1S518	202.19	AMD30	2.51	.05	Klein et al. 1998, Weeks et al. 2001
Chromosome 2:					
D2S410	125.18	AMD133	2.10	.10	
D2S1353	164.51	AMD159	1.52	.61	
D2S2944	210.43	AMD32	2.03	.25	
Chromosome 3:					
D3S4545	26.25	AMD159	1.86	.23	Iyengar et al. 2002
D3S4529	112.42	AMD7	1.78	.56	
D3S2460	134.64	AMD187	1.71	.07	
D3S2418	215.84	AMD104	1.55	.49	
Chromosome 4:					
D4S1644	143.31	AMD4	2.17	.23	
Chromosome 9:					
D9S1122/D9S922	75.88	AMD159	1.53	.60	Weeks et al. 2000
Chromosome 10:					
D10S1230	142.78	AMD104	1.55	.49	Weeks et al. 2001, Haines et al. 2002, Iyengar et al. 2002
Chromosome 12:					
D12S1064	95.03	AMD7	1.90	.33	Iyengar et al. 2002
Chromosome 17:					
D17S2193	89.32	AMD4	1.97	.29	
Chromosome 18:					
D18S844	116.44	AMD33	1.94	.21	
Chromosome 19:					
D19S591	9.84	AMD7	1.85	.48	

NOTE.—The reported *P* values are empirical genomewide false-positive rates based on 100 replicates of simulated null data.

posed of affected members with an average age of <75 years (14 families) or ⩾75 years (56 families).

For the multipoint analyses performed using Allegro (Gudbjartsson et al. 2000), nine of the families that we studied exceeded the size limit of the program and had to be trimmed. This could be done with a small loss of power, by removing unclassified or young unaffected individuals (who may still develop the disease at a later age). One large family, AMD126 (56 individuals), had to be split into two equal-size pedigrees.

Allele frequencies were estimated from the data by using PedManager. The estimates were based on the counting of allele frequencies in all genotyped individuals. Although such an approach is not as accurate as maximum-likelihood estimates that consider relationships within pedigrees, it is favorable because of its simplicity and is generally conservative with respect to final statistical significance of linkage analysis (Weeks et al. 2001). However, to confirm that our most significant signals were not due to misspecification of allele frequencies, for each of the peaks (1q25-31, 3p13, 4q32, 9q33, and 10q26), we used the maximum-likelihood estimation implemented in Ilink (Terwilliger and Ott 1994) to estimate allele frequencies for nine markers surrounding each peak. Linkage analyses were then repeated using the maximum-likelihood frequency estimates. None of the results were significantly altered; all LOD scores obtained with maximum-likelihood estimates were within 2% of the original values.

We used simulations to determine the empirical significance levels of our results. Unfortunately, because of constraints on computational time, we could analyze simulated data only for individual families (table 2), rather than for the entire data set. For each family, we used Allegro to simulate 100 replicates of null data of the entire genome scan. Genotypes were simulated in accordance with the original pedigree structures and only for the individuals who were available for genotyping in our data set. Allele frequencies for the simulation were obtained from the real data set. Genotyping efficiency (i.e., the fraction of genotypes that were unambiguously determined) was set to 85%, to account for missing and inconsistent genotypes, as well as differences in markers used in the two genotyping panels. This means that 15% of genotypes were randomly assigned an "unknown" value. The resulting *P* values represent genomewide significance levels (i.e., the fraction of replicates that contained LOD scores at least as extreme as those observed). The relatively low number of replicates was limited by constraints on computation-

al time. The results should be viewed with some caution but should provide a useful indication as to which families provide particularly strong evidence of linkage.

Correlation Analysis

To investigate the evidence for interactions between our candidate loci, we constructed a pairwise Pearson correlation matrix for family-specific LOD scores at the four candidate regions that were detected using our unstratified sample (table 3). For each of the 70 families, individual LOD scores at the location of each peak were calculated under the corresponding genetic model that was initially used to detect the peak (dominant with no age dependence for 1q25-31, 3p13, and 10q26; allele sharing for 9q33). The *P* values associated with the correlation coefficients were corrected for multiple testing by dividing by the total number of comparisons (Bonferroni correction). We used two methods to correct for departures from normal distribution of the data: (1) outlier correction (removing families with extreme values that have a large effect on the significance of analysis), performed using the Hadi outlier method implemented in Systat 8.0 software (SPSS); and (2) the nonparametric Kendall's τ correlation test.

To further exclude the possibility that the observed correlation between LOD scores may be caused by variation in family sizes and, hence, variation in information content across families, we performed the following partial correlation analysis: (1) we determined the maximum possible LOD score (LOD_{max}) for each family in the data set, by calculating the LOD score between the disease marker and itself (the LOD_{max} values represent a measure of information content); (2) we log-transformed LOD_{max} values to ensure a better approximation to a normal distribution; and (3) we used the transformed LOD_{max} value as a conditional variable, to compute the partial correlation between LOD scores observed at candidate loci. This analysis corrects the correlation for the effect of variation in family size.

Results

Linkage Analysis

Our initial analysis of 21 families uncovered a single region with LOD scores suggestive of genetic linkage. We found a LOD score of 2.07, based on the allele-sharing exponential pairs Z_{lr} statistic, on chromosome 9, between D9S930 and D9S934. After genotyping six additional markers between D9S938 and D9S1825, we found no change in the LOD scores and no change in the location of the peak. There were no other peaks exceeding a LOD-score cutoff of 1.5 in the initial genome scan.

Subsequently, we included additional families and individuals to increase the sample size, and we performed analyses of combined data by using different genetic models and diagnostic schemes (see table 1 and the "Families and Methods" section). The results are shown graphically in figure 1. Below, we describe regions with scores exceeding the LOD = 1.5 threshold under at least one analysis method. Incidentally, all of the loci reported below also exceed the LOD = 2 cutoff when we include results from the groups stratified by predominant phenotype and by age of affected members at ascertainment. A threshold of LOD = 2 is often used as a criterion for suggestive linkage. However, note that, because multiple analyses have been performed, the actual statistical significance level is lower than it would be for a single test.

Table 3

Pearson Correlation Values (ρ) and Associated *P* Values between Candidate Loci

	1q25-31	3p13	9q33	10q26
1q25-31	...	.02	.85	.15
3p13	.2805	...	.52	.002[a]
9q33	.0235	−.0794	...	.03
10q26	.1743	.3666	−.2523	...

NOTE.—Correlation values are given in the lower left of the matrix, and associated *P* values are given in the upper right of the matrix.

[a] Significant at the Bonferroni-corrected $P < .05$ level.

The highest overall peak, with an HLOD score of 3.06, was found on chromosome 10, at D10S1230 (genetic location 143 cM), using genetic model 2 and breaking down the pedigrees into nuclear families (see scheme C, in table 1 and the "Families and Methods" section). This locus has been reported previously (Weeks et al. 2001).

The second-highest peak, LOD = 2.01, was found on chromosome 9, at D9S934 (136 cM), 4 cM telomeric of the peak location on chromosome 9 in the initial 21 families. This result was obtained using the allele-sharing approach in entire pedigrees (model 3A). Note that parametric analysis of the entire data set under the age-dependent–penetrance model (model 1A) yielded an HLOD score of 1.48 at the same location. This appears to be a novel locus, not seen in any previous studies to our knowledge.

The third peak is an HLOD score of 1.81 on chromosome 3, between D3S1304 and D3S4545 (27 cM). This result was obtained using the parametric model with age-independent penetrance (model 2B). This locus has been observed in a previous study (Iyengar et al. 2002).

Our final peak in the unstratified data set, with a maximum HLOD score of 1.70 at D1S518 under parametric model 2B, corresponds to the 1q25-31 locus reported earlier by Klein et al. (1998) and Weeks et al. (2001).

To dissect potential genetic factors responsible for dif-

ferences in disease manifestation, we then stratified the sample by the two major phenotypes of advanced AMD (predominantly wet [neovascular AMD] and predominantly dry [geographic atrophy]) and by age of affected family members at the time of our evaluation. The most notable result of this analysis was an HLOD score of 2.66 at D4S2368 (on chromosome 4, at 168 cM) for the predominantly-dry-AMD subgroup under model 1A. This peak represents another novel locus, not implicated in any earlier studies to our knowledge.

Within the age-stratified sample, we found that our peak on 3p13 increased to HLOD = 2.19 (model 2A) for the sample of families with early age at ascertainment. We also found an HLOD score of 1.57 (model 2B) and $LOD_{Z|r}$ of 2.07 at 1q25-31 (marker D1S518).

We also report regions with LOD scores exceeding 1.5 in individual families. Since the age-dependent–penetrance model generally produces very low individual scores, we report scores by using the constant-penetrance LOD-score analysis of entire families (model 2A). The results are shown in table 2. Note that several of the loci map to regions that have been suggested by other studies (Klein et al. 1998; Weeks et al. 2000, 2001; Iyengar et al. 2002).

Correlation Analysis

For the four candidate regions (1q25-31, 3p13, 9q33, and 10q26) that were detected within the unstratified data set, we performed a correlation analysis (similar to that performed by Cox et al. (1999) on non–insulin-dependent diabetes mellitus data) to detect correlations between LOD scores at the peak locations within individual families. The results are shown in table 3. Our data provide evidence for a statistically significant interaction between 10q26 and 3p13 ($\rho = 0.367$; $P = .002$). After correction for multiple testing (Bonferroni correction), the correlations between the remaining pairs of loci were not statistically significant, but the Bartlett χ^2 test for sphericity of the determinant, which detects the presence of correlations within the entire table, was highly significant ($\chi^2 = 19.989$; df = 6; $P = .003$), indicating that more than one interaction may be present. Since both the Bartlett χ^2 test and the Pearson correlation analysis are sensitive to deviations of the data from normality, we performed two additional tests—Pearson correlation with outlier correction and the nonparametric Kendall's τ correlation (see the "Families and Methods" section). The correlation is robust with respect to the outlier correction ($\rho = 0.373$; $P = .001$) and remains significant under the nonparametric Kendall's τ correlation test ($r = 0.22$; $P = .008$). Finally, to exclude the possibility that the observed correlation may be caused by variation of information content across families of different sizes, we performed partial Pearson correlation analysis, correcting for the effect of family size (see the "Families and Methods" section). Again, the partial coefficient of correlation between 10q26 and 3p13 is significant ($\rho = 0.334$; $P = .005$). All of the above tests of the correlation between 10q26 and 3p13 are significant after correction for multiple testing within the 4 × 4 correlation matrix.

Discussion

AMD is the leading cause of vision loss in elderly Americans. As such, it has recently acquired close attention from geneticists. Previously, our group conducted linkage analysis in a large family (AMD30) and identified a candidate locus on 1q25-31. The segregation of the disease within family AMD30 was largely in agreement with a dominant Mendelian mode of inheritance. Since then, one group has published the results of their genomewide scan (Weeks et al. 2000, 2001), and another group has reported preliminary data on large collections of affected families (Iyengar et al. 2002). Neither groups' results indicated the presence of a single, major AMD locus, and, with one exception (Weeks et al. 2001), suggestive peaks were generally low (LOD < 3). Thus, it is likely that, whereas AMD is caused by a major effect of a single gene in some pedigrees (e.g., AMD30), the disease may have more complex underlying genetic mechanisms in other pedigrees. There may be a high level of genetic locus heterogeneity, resulting in different loci segregating with the disease in various families. Moreover, there may exist complex epistatic interactions between the loci, with interactions between different sets of loci resulting in similar phenotypes, even in related individuals.

Despite the availability of large sample sizes, LOD scores obtained by various groups failed to provide clear evidence of linkage at the commonly accepted statistical significance levels, particularly after accounting for the multiple diagnosis and analysis methods commonly employed in each study. In view of the above problems, emphasis must be placed on cross-validation of results obtained by different groups, even without individually achieving statistical significance, as a method to narrow the regions of interest. In addition, we find that the uncertain mode of inheritance of AMD justifies and necessitates the use of several analysis methods. Although the use of multiple models decreases the actual significance levels (and hence increases false-positive rates) of the results, it remains a valid approach to the identification of the most likely candidate regions. Our present analysis provides evidence in support of candidate loci suggested by earlier studies, as well as evidence of new candidates that should be taken into consideration in future linkage analyses.

First, let us consider the 1q25-31 locus, which we

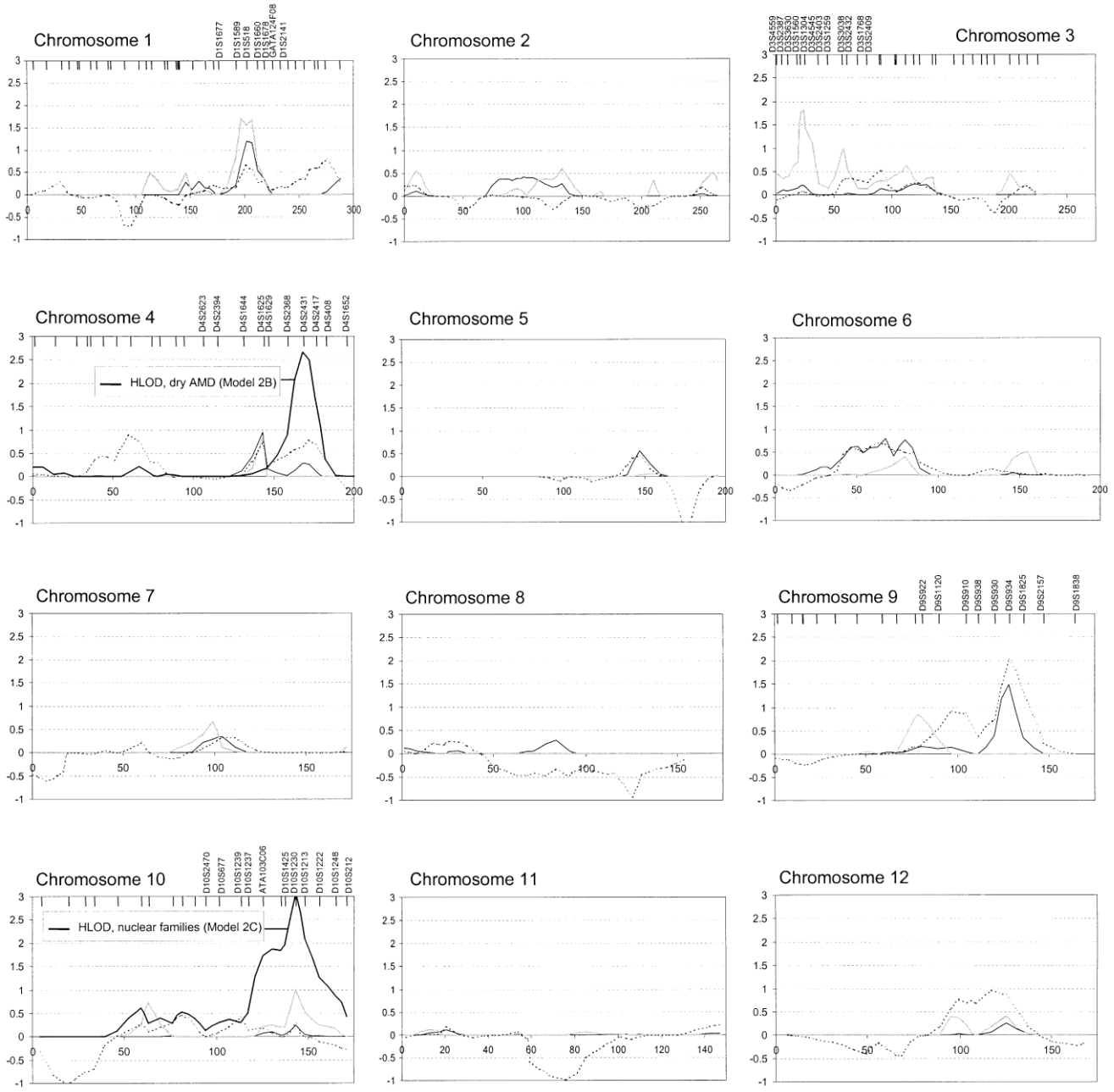
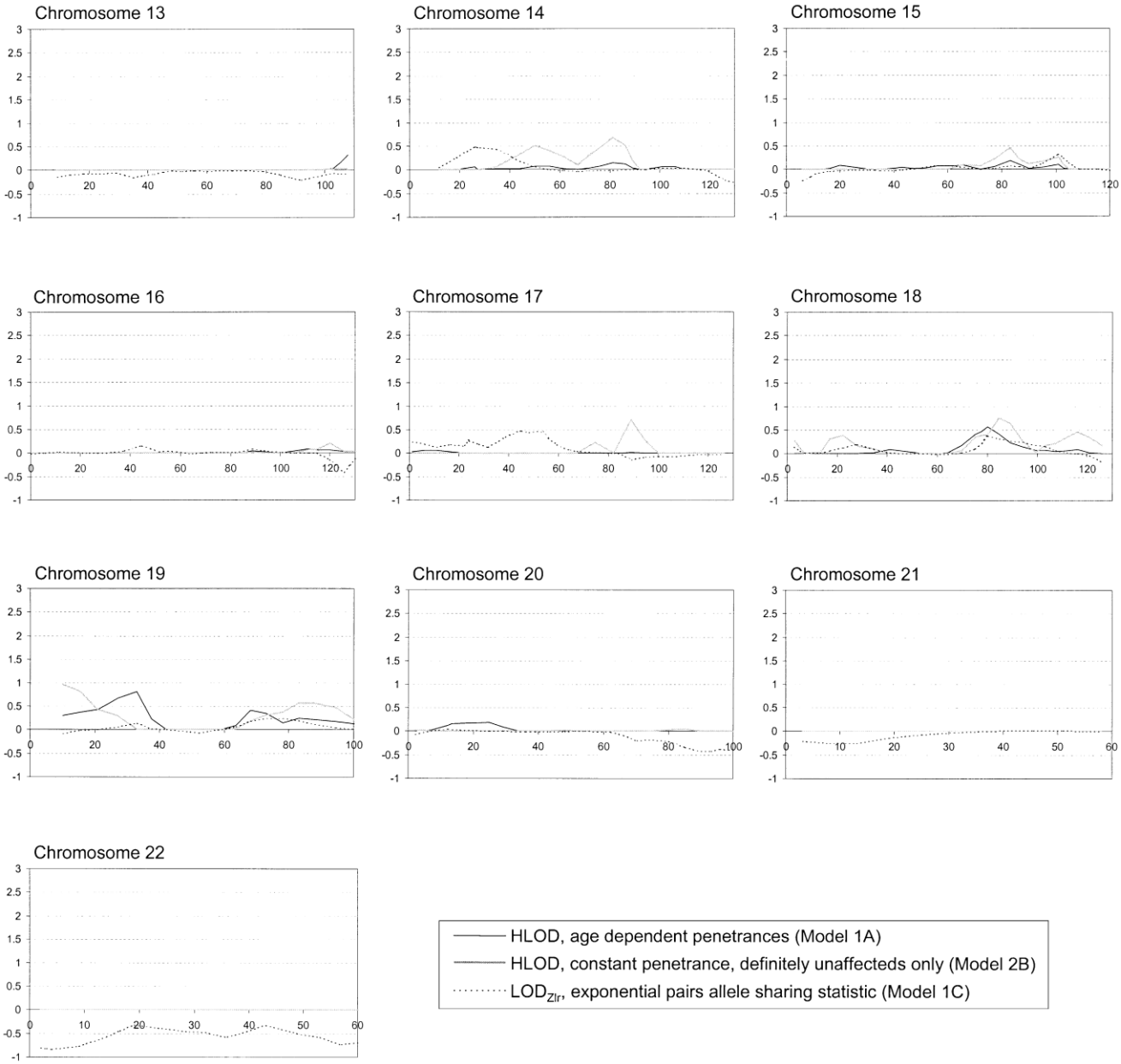


Figure 1 Genomewide linkage results of the three most significant analyses (models 1A, 2B, and 3A [for summary, see table 1]). Chromosomes 4 and 10 data include additional analyses (under models 2B-dry and 2C, respectively) that have produced notable results. The candidate regions, on chromosomes 4 and 9, constitute novel loci and have not been reported in previous studies. The remaining regions, on chromosomes 1, 3, and 10, have been also reported by previous and ongoing studies. The genetic locations correspond to the Marshfield genetic map (Broman et al. 1998) and are expressed in Kosambi cM. Note that the HLOD and $LOD_{Z_{lr}}$ curves represent statistics with different numbers of degrees of freedom (for a brief discussion regarding the comparison of the two statistics, see the "Families and Methods" section).

previously identified with a LOD score of 3 in one large family, AMD30, and which Weeks et al. (2001) confirmed with an HLOD score >2. In our extended sample, we see a maximum-HLOD-score peak of 1.70 under genetic model 2B. We find one additional large family, AMD159, consisting of 25 individuals (14 of whom were genotyped) and including 6 affected and 5 definitely unaffected individuals, that shows evidence of linkage at this locus (LOD = 2.59, under model 2B). All of the haplotypes within this family were consistent



with linkage, providing further evidence for the existence of an AMD gene on 1q25-31.

The analysis performed by Weeks et al. (2001), however, has suggested that the percentage of families with AMD that are segregating a gene at 1q25-31 is very high, >40%. In our sample, depending on the model used, the maximum-likelihood estimate for the proportion of linked families varies between 0.07 and 0.15. Hence, although our data provide additional support for linkage to 1q25-31, we believe that the proportion of AMD cases that can be explained by this locus is considerably lower than had been previously suggested and is probably <15%. Note that the difference between the two studies may be partially due to the larger size of the families that we studied and that the interpretation of estimates of proportions of linked families may be problematic and should generally be treated with caution (Whittemore and Halpern 2001).

The locus on 10q26 (HLOD = 3.06, under model 2C) is our most significant candidate. It has also been implicated in at least two other independent AMD linkage studies. Weeks et al. (2000, 2001) found a maximum nonparametric linkage (NPL) score of 1.42 in their initial study and an S_{all} LOD score of 2.02 in their expanded sample, at D10S1230. Iyengar et al. (2002) used allele-sharing methods to obtain a *P* value of .023 at D10S212 (20 cM telomeric of D1S1230). Haines et al. (2002) found a maximum LOD score (MLS) of 1.12 at D10S1230, the highest peak of the four candidate regions that they investigated.

The detection of this locus in all of the studies performed to date suggests that 10q26 may contain a common AMD gene, either a causative (major effect) gene or a modifier (a gene that controls the effects of the major locus, e.g., mode of inheritance, penetrance, and age at onset). Interestingly, a modifier gene for the age at onset of two neurodegenerative disorders, Alzheimer disease and Parkinson disease, has recently been suggested to reside on 10q (Li et al. 2002) and may correspond to the same locus. Furthermore, we propose that the locus at 10q26 may act through a frequent allele. Our data support this hypothesis, since the analysis becomes significant only after subdivision of our extended pedigrees (maximum HLOD = 1.10) into nuclear families (HLOD = 3.06). If an allele is common, it may enter extended pedigrees through multiple married-in individuals, rendering parametric linkage analysis ineffective (Badner et al. 1998). However, subdivision of the families ensures that, within each nuclear family, the disease allele may enter the pedigree only through the parents, rendering multiple origins less likely and linkage analysis more effective.

The peak on 9q33 is our most robust peak with respect to the analysis method. It was observed in our initial sample, in the expanded sample, and in both parametric and nonparametric analyses, as well as after subdivision of the pedigrees into nuclear families. Although Weeks et al. (2000) mention a possible AMD locus at D9S1838, we believe our 9q33 locus to be novel, because our peak at D9S934 is ~38 cM proximal to their peak at D9S1838; furthermore, although our LOD score at 9q33 remained relatively unchanged when additional families were genotyped, their LOD score at the D9S1838 locus decreased significantly on doubling the number of sib pairs analyzed and was no longer mentioned as a possible AMD locus (Weeks et al. 2001). We suggest that 9q33 should be considered as a candidate locus in future screening studies.

Our peak on 3p13 has been described in the study by Iyengar et al. (2002) ($P = .012$). This peak occurs only in one type of analysis in the present study—namely parametric model 2B without age-dependent penetrance, with only definitely unaffected individuals included. Once again, we varied the model parameters and sample content to find the conditions under which this peak is maximized. We found the most significant factor to be the type of families included in the sample. When we limited the sample to pedigrees in which the affected individuals had predominantly (>50%) dry AMD, the HLOD score at 3p13 increased to 2.19.

Within the combined collection of families in the present study, we found no evidence supporting linkage to two other loci indicated in other studies, namely 9p13 and 17q25 (Weeks et al. 2000, 2001). We found slight evidence supporting another candidate locus, 12q22 (Weeks et al. 2000; Iyengar et al. 2002). Our most significant peak close to this locus was a Z_{lr}-based LOD score of 0.96 between the markers PAH and D12S2070 (at 113.5 cM).

We also investigated maximum-LOD-score signals within individual families. This approach should be effective in view of the apparent genetic heterogeneity of AMD. However, with the exception of families AMD30 and AMD159, we found that most families failed to reach their maximum possible LOD scores (in which case all haplotypes would be consistent with linkage and there would be no recombination between the candidate region and the disease locus). We find two possible explanations: First, although most investigators consider neovascular and atrophic forms of AMD as different manifestations of the same disorder, mutation of some genes may predispose the individual to a disease dominated by only one of the two phenotypes; thus, the definition of AMD used in the present study may encompass different underlying genetic mechanisms. The second explanation invokes both the high genetic heterogeneity of the disease—multiple disease loci segregating even within individual families—as well as complex epistatic interactions between loci. Note that several of the peaks observed in individual families correspond to loci indicated by prior studies—specifically, D3S4545, on 3p (Iyengar et al. 2002); D9S1122/D9S922, on chromosome 9 (close to the minor peak observed by Weeks et al. [2001]); D10S1230, on chromosome 10 (Weeks et al. 2000, 2001; Haines et al. 2002; Iyengar et al. 2002); and D12S1064, on chromosome 12 (within 10 cM of the most significant peak observed by Iyengar et al. [2002]). The remaining loci in table 2 have not been observed before. All the candidate loci will be fine-mapped and screened for candidate genes in the later stages of this project.

In the analysis of individual families, we find that several families have peaks exceeding LOD = 1.5 at several different loci—including families AMD4 (chromosomes 4 and 17), AMD7 (chromosomes 3, 12, and 19), AMD104 (chromosomes 3 and 10), and AMD159 (chromosomes 1, 2, 3, and 9)—lending additional sup-

port to the hypothesis of cosegregation of distinct loci and, hence, to a complex basis of the disease. To test the above hypothesis, we performed a correlation analysis between family-specific LOD scores at the four loci observed in our unstratified sample: 1q25-31, 3p13, 9q33, and 10q26. We found a strong indication of an overall presence of interactions ($P = .003$). The effect was mostly due to a correlation between 3p13 and 10q26 ($P = .002$). A positive correlation coefficient may be interpreted as evidence of an epistatic interaction, whereas a negative correlation may also represent genetic heterogeneity (Cox et al. 1999; Majewski et al. 2001). The 3p13-by-10q26 interaction is associated with a positive correlation coefficient, supporting both the hypothesis that AMD is a complex genetic disorder and our suggestion that 10q26 may harbor a common modifier locus (see above).

In conclusion, we have found evidence of new loci contributing to AMD, at 4q32 and 9q33. These loci will be studied in future mapping and mutational analyses. We also found additional evidence supporting loci indicated by other studies—in particular, 1q31, 10q26, and 3p13—and indication of epistatic interactions between candidate loci. The discovery of the genes at these loci and the study of the products encoded will be important future steps in the unraveling of the processes that result in the common phenotypes of AMD. In view of the complex nature of the genetics of AMD, we believe that all of these loci have the potential to harbor genes that cause AMD, either alone or in combination, or genes that act as modifiers to the age at onset, degree of severity, or eventual phenotype. Many of the gene products may be found to interact in common, if not novel, cellular functions and biochemical pathways that have yet to be discovered. The understanding of the molecular biology and biochemistry that underlie AMD will allow the creation of molecularly based pharmaceuticals and will help in the identification of new therapeutic interventions to prevent or slow the course of these blinding disorders.

Acknowledgments

This work was supported by grant HG00008 (to J.O., T.C.M., and J.M.) National Human Genome Research Institute, Bethesda, MD; grant EY 12203 (to M.L.K.) and grants EY 03279, EY 08247, and EY 10572 (all to T.S.A.) from the National Institutes of Health, Bethesda, MD; the Collins Medical Trust, Portland, OR (support to D.W.S.); The Foundation Fighting Blindness, Owings Mills, MD (support to D.W.S. and R.G.W.); the George and Carolyn Goodall Macular Degeneration Fund; and an unrestricted grant from Research to Prevent Blindness, New York. We thank Blair Berselli, Susan Nolte, Heather Cooper, and Genet Friess, for help with family ascertainment, and Rebecca Barra, for sample DNA amplification.

Electronic-Database Information

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for AMD)

References

Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, Zabriskie NA, Li Y, Hutchinson A, Dean M, Lupski JR, Leppert M (1997) Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. Science 277:1805–1807

Badner JA, Gershon ES, Goldin LR (1998) Optimal ascertainment strategies to detect linkage to common disease alleles. Am J Hum Genet 63:880–888

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Cottingham RW Jr, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A (1999) Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nat Genet 21:213–215

Evans J, Wormald R (1996) Is the incidence of registrable age-related macular degeneration increasing? Br J Ophthalmol 80:9–14

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Haines JL, Lamb JM, Schmidt S, Agarwal A, Postel E, De La Paz-Schmidlekofer M, Gilbert JR, Scott WK, Pericak-Vance MA (2002) Follow-up of genomic screen regions in age-related macular degeneration. Paper presented at Association for Research in Vision and Ophthalmology Meeting, Fort Lauderdale, FL, May 5–10

Iyengar SK, Schick J, Reading K, Milliard C, Brey N, Liptak R, Klein R, Klein B, Elston R (2002) A genomewide scan for age-related maculopathy and other ocular phenotypes in a sample from the Beaver Dam Eye Study (BDES). Paper presented at Association for Research in Vision and Ophthalmology Meeting, Fort Lauderdale, FL, May 5–10

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The Human Genome Browser at UCSC. Genome Res 12:996–1006

Klein ML, Schultz DW, Edwards A, Matise TC, Rust K, Berselli CB, Trzupek K, Weleber RG, Ott J, Wirtz MK, Acott TS (1998) Age-related macular degeneration: clinical features in a large family and linkage to chromosome 1q. Arch Ophthalmol 116:1082–1088

Klein R, Klein BE, Linton KL (1992) Prevalence of age-related maculopathy: the Beaver Dam Eye Study. Ophthalmology 99: 933–943

Klein R, Klein BE, Tomany SC, Meuer SM, Huang GH (2002) Ten-year incidence and progression of age-related maculopa-

thy: the Beaver Dam Eye Study. Ophthalmology 109:1767–1779

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, et al (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. Am J Hum Genet 70:985–993

Majewski J, Li H, Ott J (2001) The Ising model in physics and statistical genetics. Am J Hum Genet 69:853–862

McPeek MS (1999) Optimal allele-sharing statistics for genetic mapping using affected relatives. Genet Epidemiol 16: 225–249

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Ott J (1999) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

Terwilliger J, Ott J (1994) Handbook of human genetic linkage. Johns Hopkins University Press, Baltimore

Weeks DE, Conley YP, Mah TS, Paul TO, Morse L, Ngo-Chang J, Dailey JP, Ferrell RE, Gorin MB (2000) A full genome scan for age-related maculopathy. Hum Mol Genet 9:1329–1349

Weeks DE, Conley YP, Tsai HJ, Mah TS, Rosenfeld PJ, Paul TO, Eller AW, Morse LS, Dailey JP, Ferrell RE, Gorin MB (2001) Age-related maculopathy: an expanded genomewide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. Am J Ophthalmol 132:682–692

Whittemore AS, Halpern J (2001) Problems in the definition, interpretation, and evaluation of genetic heterogeneity. Am J Hum Genet 68:457–465

Williamson JA, Amos CI (1995) Guess LOD approach: sufficient conditions for robustness. Genet Epidemiol 12:163–176