An Analysis of Allelic Variation in the *ABCA4* Gene

Andrew R. Webster,[1,12] *Elise Héon,*[2,3] *Andrew J. Lotery,*[1] *Kimberlie Vandenburgh,*[1] *Thomas L. Casavant,*[4] *Kean T. Oh,*[1] *Gretel Beck,*[5] *Gerald A. Fishman,*[6] *Byron L. Lam,*[7] *Alex Levin,*[3] *John R. Heckenlively,*[8] *Samuel G. Jacobson,*[9] *Richard G. Weleber,*[10] *Val C. Sheffield,*[6,11] *and Edwin M. Stone*[1]

Purpose. To assess the allelic variation of the ATP-binding transporter protein (*ABCA4*).

Methods. A combination of single-strand conformation polymorphism (SSCP) and automated DNA sequencing was used to systematically screen this gene for sequence variations in 374 unrelated probands with a clinical diagnosis of Stargardt disease, 182 patients with age-related macular degeneration (AMD), and 96 normal subjects.

Results. There was no significant difference in the proportion of any single variant or class of variant between the control and AMD groups. In contrast, truncating variants, amino acid substitutions, synonymous codon changes, and intronic variants were significantly enriched in patients with Stargardt disease when compared with their presence in subjects without Stargardt disease (Kruskal–Wallis $P < 0.0001$ for each variant group). Overall, there were 2480 instances of 213 different variants in the *ABCA4* gene, including 589 instances of 97 amino acid substitutions, and 45 instances of 33 truncating variants.

Conclusions. Of the 97 amino acid substitutions, 11 occurred at a frequency that made them unlikely to be high-penetrance recessive disease-causing variants (HPRDCV). After accounting for variants in *cis*, one or more changes that were compatible with HPRDCV were found on 35% of all Stargardt-associated alleles overall. The nucleotide diversity of the *ABCA4* coding region, a collective measure of the number and prevalence of polymorphic sites in a region of DNA, was found to be 1.28, a value that is 9 to 400 times greater than that of two other macular disease genes that were examined in a similar fashion (*VMD2* and *EFEMP1*). (*Invest Ophthalmol Vis Sci.* 2001;42:1179–1189)

The ABCA4 protein, a member of the superfamily of ABC (ATP-binding cassette) transporter proteins, is thought to transport vitamin A derivatives across the disc membrane.[1,2] Initial studies on primate retina suggested exclusive expression in retinal rod photoreceptors.[3,4] However, recent immunohistochemical studies on human retina suggest that expression also occurs within macular cones.[5] ABC transporter proteins exist in both prokaryotic and eukaryotic organisms. They couple the hydrolysis of the gamma phosphodiesterase bond of adenosine triphosphate (ATP) to the translocation of a substrate across a lipid bilayer. Each ABC transporter protein contains two regions of homology in its nucleotide-binding domain (A and B Walker motifs,[6] as well as an additional C motif that distinguishes this class of ATP binding proteins from kinases, myosins, and ATP-synthases).[7,8] Variation in these proteins cause a variety of important human diseases, including cystic fibrosis,[9] progressive familial intrahepatic cholestasis,[10] and the *ABCA4*-associated diseases Stargardt macular dystrophy, cone–rod dystrophy, and retinitis pigmentosa.[3,11,12]

Stargardt disease, or fundus flavimaculatus, is a recessively inherited macular dystrophy estimated to affect 1 in 10,000 people.[13] Typically, the disorder is characterized by difficulty with central vision that eventually progresses to the level of legal blindness. The peripheral vision is usually preserved, even late into the course of the disease. There is a characteristic "beaten bronze" appearance of the central region of the retina (the macula), with small yellow flecks scattered more peripherally. In the majority of affected patients, fluorescein angiography reveals a masking of the fluorescence of the choroidal circulation ("dark choroid") suggesting a widespread abnormality at the level of the retinal pigment epithelium.[14]

In 1994, Stargardt disease was mapped to chromosome 1q[15] and was later shown to be due to variations in a gene now known as *ABCA4*.[3] The gene for an autosomal dominant disorder with a similar phenotype has been reported on chromosome 6.[16] Autosomal dominant families linked to this locus are at least 50 times less common than families that are consistent with autosomal recessive inheritance.[17,18] Recessively inherited Stargardt disease is likely to be monogenic; no sibships have been found in which *ABCA4* alleles do not cosegregate with the disease. However, intrafamilial variability of the phenotype is commonly reported.[19–21] Some siblings can exhibit purely macular disease, whereas others have only peripheral flecks.[22] This suggests that other genes in the genetic background play a role in the expression of a given *ABCA4* genotype. In addition to Stargardt disease, a number of families have been reported in which other retinal phenotypes such as cone–rod dystrophy and retinitis pigmentosa are associated with *ABCA4* sequence variations. Some families with the latter phenotype harbor *ABCA4* variants that would be expected to entirely eliminate *ABCA4* function.[11,12,23]

From the Departments of [1]Ophthalmology and [6]Pediatrics, The University of Iowa College of Medicine, Iowa City; [2]Institute of Ophthalmology and Moorfields Eye Hospital, London, United Kingdom; the [3]Department of Ophthalmology, University of Toronto, Canada; [4]Vision Science Research Program, University Health Network, Toronto, Canada; the [5]Department of Electrical Engineering and Computer Science, The University of Iowa, Iowa City; [7]University of Illinois Eye and Ear Infirmary, Chicago; [8]Bascom Palmer Eye Institute, University of Miami, Florida; [9]The Jules Stein Eye Institute, The University of California at Los Angeles; the [10]Department of Ophthalmology, Scheie Eye Institute, Philadelphia, Pennsylvania; [11]The Casey Eye Institute, The Oregon Health Sciences University, Portland; and [12]The Howard Hughes Medical Institute, Iowa City.

Supported in part by Grant EY10539 from the National Institutes of Health; The Foundation Fighting Blindness; The Carver Charitable Trust; The T. F. C. Frost Trust; The Ronald McDonald House Charities; The Macula Vision Research Foundation; The Grousbeck Family Foundation; The Wellcome Trust, United Kingdom; and an unrestricted grant from Research to Prevent Blindness. AJL is a recipient of a Research to Prevent Blindness Career Development Award.

Submitted for publication September 26, 2000; revised January 3, 2001; accepted January 18, 2001.

Commercial relationships policy: N.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Edwin M. Stone, Department of Ophthalmology, The University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242. edwin-stone@uiowa.edu

Copyright © Association for Research in Vision and Ophthalmology

In this article, we present a comprehensive survey of the entire exonic and flanking intron sequence of the *ABCA4* gene in 374 unrelated patients with a clinical diagnosis of Stargardt disease, 182 patients with age-related macular degeneration (AMD), and 96 normal subjects from the same clinic population as the patients with AMD. We compared the variants we observed with the mouse and bovine *ABCA4* sequences and with the sequences of other members of the ABC superfamily, to help predict the clinical significance of the missense changes we observed. Finally, to quantify the degree of polymorphism of the *ABCA4* gene in our population, we used a statistic known as nucleotide diversity, to compare *ABCA4* with two other human macular disease genes (*VMD2,* Best disease, and *EFEMP1*, Malattia Leventinese–Doyne honeycomb retinal dystrophies) that we have recently examined in a similar fashion.[24,25]

METHODS

Recruitment of Subjects

The protocol of the study conformed to the Declaration of Helsinki, and informed consent was obtained from all subjects. Patients with Stargardt disease were recruited from the retina clinic of the University of Iowa as well as from other retina specialists from throughout the United States and Canada. All probands included in this study had been previously screened for variation in the coding sequence of the *RDS* gene and found to be normal. All patients exhibited clinical features that were thought to be consistent with a diagnosis of Stargardt disease by their referring physicians. A subset of all probands who had been personally examined by at least one of the authors and were judged to have characteristic features of Stargardt disease were analyzed separately and compared with the Stargardt cohort as a whole. Unrelated patients with AMD were recruited exclusively from the University of Iowa. Normal subjects were also drawn from the latter clinic population and had an ethnic distribution that was very similar to that of the patients with AMD. All these normal subjects were more than 40 years of age, had normal maculae and intraocular pressures, and had no family history of Stargardt disease. The AMD group included 109 (59.9%) patients in whom, by the time of the study, a choroidal neovascular (CNV) membrane or a pigment epithelial detachment without a detectable CNV had developed in at least one eye and 73 (40.1%) who had not had such a complication by the time of ascertainment in either eye. All 182 patients with AMD had a degree of macular disease in one or both eyes sufficient to be designated AMD on the basis of the International Classification of the International ARM and Epidemiologic Study Group.[26]

Molecular Methods

DNA was extracted from venous blood as previously described. All 50 exons of the *ABCA4* gene were amplified using 51 primer pairs as previously described.[3,27] For single-strand conformation polymorphism (SSCP) analysis, amplimers were denatured at 95°C for 3 minutes, electrophoresed on 6% polyacrylamide–5% glycerol nondenaturing gels, and stained with silver nitrate.[28,29] Samples exhibiting mobility shifts were sequenced using fluorophore-labeled dye-terminator chemistry on a sequencer (model 377; Perkin Elmer–ABI, Foster City, CA), and the sequences were analyzed using commercially available software (Sequencher; Gene Codes, Ann Arbor, MI). To examine the sensitivity of SSCP with respect to direct sequencing, the *ABCA4* coding regions of seven probands, each randomly selected from the group of patients who exhibited one likely high-penetrance recessive disease-causing variant (HPRDCV; described in the next section) in the conventional assay, were bidirectionally sequenced.

Definitions

In this article, the term variant is used to refer to one or more contiguous nucleotides that differ from the most common sequence in the population. Theoretically, these variants can be divided into those that never under any circumstances alter the phenotype of an individual (non–disease-causing variants; NDCVs) and those that can alter the phenotype of an individual who harbors them (disease-causing variants; DCVs). The latter class can be further subdivided into those variants that would be expected to alter the phenotype in a high proportion (e.g., 90%) of individuals who carry them in the appropriate configuration (high-penetrance DCVs [HPDCVs]) and those variants that would alter the phenotype of the individual only in combination with other factors, such as other elements in the genetic background (low-penetrance DCVs). The two categories of DCVs can be further subdivided into those that are expected to alter the phenotype in the heterozygous state (e.g., high-penetrance dominant disease-causing variants; HPDDCVs) and those that would be expected to alter only the phenotype in the homozygous or compound heterozygous state (e.g., HPRDCVs).

Sequence variants that lie outside the *ABCA4* gene and whose presence increase or decrease the likelihood that an altered phenotype will result from the presence of a DCV within *ABCA4* are referred to as threshold-modifying variants.

The term allele is used to refer to all variants present in a single copy of a gene. Because in some cases it is the combination of these variants that actually gives rise to the phenotype, there is a parallel nomenclature for alleles to that just described for variants. For example, a high-penetrance recessive disease-causing allele (HPRDCA) refers to a group of one or more variants acting in *cis* that have a high probability of altering the phenotype of an individual who harbors them when they are present in the homozygous or compound heterozygous state. Because some alleles consist of single variants, it may not be obvious why the allele nomenclature cannot be used in all cases. The reason is that correct detection of the phase of the variants is necessary for the proper recognition of alleles, and this detection often requires genotypic information from parents or other relatives that is not always available. In contrast, the presence or absence of individual variants can be correctly detected in individual study participants.

Although the foregoing definitions provide a useful frame of reference for discussions about the relationship between certain genotypes and phenotypes, it is important to recognize that in practice it can be quite difficult to assign a specific variant to one of these theoretical categories. In the absence of an accurate assay of a gene's function in humans, the functional consequences of a variant in a human being can be inferred, at best, by observing its association with certain phenotypes in a large number of individuals and by judiciously using information gleaned from the study of other genes to aid in the interpretation. For example, although it is possible for a synonymous codon change in a gene to result in alteration of a phenotype,[30] it is far more common for such changes to have no effect on the phenotype. Thus, in this study, synonymous codon changes are assumed to be NDCVs, even though they are enriched in patients with Stargardt disease. Similarly, sequence variations that cause significant changes in the structure of the encoded protein are often found, as a group, to be more likely to be associated with disease than are changes with lesser effects on the encoded protein. In this article, we use the term nonconservative sequence variant to refer to one that is predicted to cause premature protein termination, aberrant splicing, an in-frame deletion or insertion, or a change in amino acid charge, size, or polarity. This definition is the same as that used previously for the comparison of *GLC1A* variants in normal subjects and subjects with glaucomatous eyes.[31]

Statistical Methods and Data Analysis

Data for all variants were entered onto a database (Access; Microsoft, Redmond, WA), and subsequent analysis was performed by computer (Access and Excel software; Microsoft). For each category of nucleotide variation, the number of variants per person was compared among groups using the Kruskal–Wallis nonparametric analysis of variance

(three groups) or the Wilcoxon rank-sum test (two groups). The proportion of subjects showing one or more specific variants or groups of variants was compared between groups using Fisher's two-tailed exact test. A 95% confidence interval for SSCP sensitivity was calculated using the binomial distribution. To determine whether any combination of polymorphisms in the *ABCA4* gene occur more commonly in patients with AMD than in control subjects, we identified the 20 most common polymorphic variants in the *ABCA4* gene and tested all pairs and all triplets of these variants in patients with AMD and in control subjects, using a program written for this purpose (Visual Basic; Microsoft). The same approach was used to determine whether a specific high-risk allele existed more commonly in patients with Stargardt disease without any detectable *ABCA4* sequence variations than in patients with Stargardt disease in whom one or both mutant alleles had been detected or in control individuals. This type of analysis has the potential to reveal the existence of a common, disease-causing Stargardt allele, whose actual DCV is undetectable by the methods used in this study. The basis of this strategy is the identification of linkage disequilibrium between the true (but undetectable) DCV and a haplotype of detectable *ABCA4* sequence variations.

The data were analyzed using two different sets of assumptions (models) to see which was more consistent with our observations. In model A, Stargardt disease was considered to be an ideal autosomal recessive disease with a prevalence of 1 in 10,000. All variant alleles were assumed to fall into one of two nonoverlapping classes: disease causing or non–disease causing. All DCVs were assumed to be equally (and highly) penetrant, so that any combination of two DCVs would be equally likely to result in the clinical phenotype of Stargardt disease. All alleles were assumed to be in Hardy–Weinberg equilibrium, and epigenetic effects of other variants in *cis* or in *trans* were assumed to be small enough to ignore. All study participants were assumed to be drawn from an ethnically identical population. The observed prevalence of each variant allele in patients with Stargardt disease and in those without was tested against two hypotheses: hypothesis one (H_1), in which the prevalence of the variant is equal in patients with Stargardt disease and in those without, and hypothesis two (H_2), in which the variant is either disease-causing or is in complete disequilibrium with a disease-causing variant, so that its prevalence in patients with Stargardt disease is 100 times that seen in those without (the expected distribution of an allele in Hardy–Weinberg equilibrium, assuming a prevalence of recessive disease of 1 in 10,000). H_1 was tested using Fisher's exact test. H_2 was tested using the binomial distribution to estimate the probability of the observed prevalence of alleles in the Stargardt and non-Stargardt groups given the null hypothesis of a 100:1 distribution in Stargardt to non-Stargardt alleles for a fully penetrant allele.

In model B, alleles were assumed to be variably pathogenic, and the clinical phenotype of Stargardt disease was assumed to result from the presence of two alleles whose combined pathogenicity fell within a fairly narrow range. Individuals with *ABCA4* alleles whose combined pathogenicity was below this range would appear clinically normal, whereas individuals whose combined pathogenicity was above this range would manifest a retinal disease that would be clinically different from Stargardt disease. To test this model, we identified variants that were enriched in patients with Stargardt disease but to a lesser extent than the expected 100:1 ratio that would be seen under model A. We then investigated whether such alleles more commonly occur in *trans* with alleles that would be expected to have a severe or null effect on the protein (which would support model B). We also investigated whether such variants occur homozygously or in *trans* with each other in patients with Stargardt disease (which would refute model B).

Sequence Comparison

Sequence comparison between organisms, between homologous ABC proteins, and between the two nucleotide-binding domains (NBDs) of ABCA4 itself, was performed using the BLAST algorithm at the National Center for Biotechnology Web site (provided in the public domain by the National Institutes of Health, Bethesda, MD, and available at http://www.ncbi.nlm.nih.gov/gorf/bl2.html).

A Statistic to Describe Nucleotide Diversity

During the study, it became evident that the general population exhibits a much greater degree of sequence variation in the *ABCA4* gene (both in total number and variety of sequence changes) than it does in other macular disease genes that we have analyzed. To quantify the degree of allelic complexity of a gene such as *ABCA4* within a specific population, we used a statistic known as nucleotide diversity.[32] This statistic quantifies the combination of two properties of a gene: (1) the number of different polymorphic sites in a gene and (2) the likelihood that two chromosomes in a population are different at these polymorphic sites. The value of the nucleotide diversity statistic is equivalent to the mean number of nucleotide positions that would be heterozygous in a pair of chromosomes after repeat samplings of a chromosomal pair from a population. The nucleotide diversity value is not dependent on the degree of linkage disequilibrium of the individual polymorphic sites. We calculated this value for the coding region of the *ABCA4* gene and compared this with the value for the *EFEMP1* and *VMD2* genes, previously analyzed in a similar fashion,[24,25] in the 96 control patients in this study. The following equation was used[32]:

$$\sum_{j=1}^{j=m}\left(1-\sum_{i=1}^{i=n}p_{ij}^2\right)$$

where p_{ij} is prevalence of the ith of n alleles at the jth of m polymorphic sites.

Results

Overall Degree of Polymorphism of the *ABCA4* Gene

In all, 2480 instances of 213 distinct nucleotide polymorphisms were detected in the *ABCA4* gene in the 652 individuals in this study (Tables 1, 2, 3). One hundred fifty-five occurred in the coding region of the gene, including 97 amino acid substitutions or deletions, 33 variants predicted to cause premature protein termination or aberrant mRNA splicing, and 24 synonymous codon changes. Fifty-eight different variable sites were detected in flanking intronic sequence and seven of these occurred in more than 4% of all alleles (Table 3). Overall, sequence variation was detected in 121 (1.4%) of the 6819 bases of the *ABCA4* open reading frame. When patients with Stargardt disease were excluded, only 27 of these positions were found to show some degree of variation (0.4%). At the protein level, variation was found in 97 (4.3%) of 2273 amino acids.

When patients with Stargardt disease were excluded, only 16 of these amino acids were found to vary (0.7%). Four amino acid substitutions were found in more than 4% of all alleles (Table 1). In non-Stargardt alleles, the proportion of bases showing some degree of polymorphism was slightly higher in intronic (27/5677) than exonic (28/7309) sequence, but this was not statistically significant (Fisher's exact test $P = 0.41$). Of the 213 variants observed, 185 were single-nucleotide substitutions. There was a significant preponderance of transitions (125/185), which probably reflects the fact that deamination of methylated cytosine residues (90/185 variants were $C \rightarrow T$ or $G \rightarrow A$ changes) is a common mechanism for point mutation.

The nucleotide diversity of the *ABCA4* coding region in non-Stargardt subjects was 1.28 sites, with a density of 1.8×10^{-4} sites per nucleotide. Therefore, a random sample of two *ABCA4* coding region sequences from our control population, when repeated many times, would, on average, show 1.28

Table 1. Thirty-Three Truncated and 98 Amino Acid–Changing Variants in the *ABCA4* Gene

Exon	Nucleotide Change	Effect	(A)	(B)	AMD (*n* = 182)	Control (*n* = 96)	STGD (*n* = 374)	Allele Prevalence
2	106delT	FS	NS		0	0	1	<0.01
2	160 + 1g → a	Splice site	NS		0	0	1	<0.01
3	161G → A	Cys54Tyr	NS		0	0	6	<0.01
3	179C → T	Ala60Val	NS		0	0	2	<0.01
3	194G → A	Gly65Glu	NS		0	0	2	<0.01
3	223T → G	Cys75Gly	NS		0	0	2	<0.01
3	247delCAAA	FS	NS		0	0	2	<0.01
3	298C → T	Ser100Pro	NS		0	0	1	<0.01
5	454C → T	Arg152Stop	NS		0	0	2	<0.01
6	574G → A	Ala192Thr	NS		0	0	1	<0.01
6	618C → G	Ser206Arg	NS		0	0	3	<0.01
6	634C → T	Arg212Cys	0.02	Yes	0	0	7	0.01
6	635G → A	Arg212His	NS		2	2	6	0.01
6	658C → T	Arg220Cys	NS		0	0	2	<0.01
6	661delG	FS	NS		0	0	1	<0.01
6	666delAAAGACGGTGCGC	FS	NS		0	0	1	<0.01
6	746A → C	Asp249Gly	NS		0	0	1	<0.01
8	899C → A	Thr300Asn	NS		0	0	1	<0.01
8	997C → T	Arg333Trp	NS		0	0	1	<0.01
9	1140T → A	Asn380Lys	NS		0	0	1	<0.01
9	1222C → T	Arg408Stop	NS		0	0	1	<0.01
10	1268A → G	His423Arg	NS		1	0	7	0.01
10	1335C → G	Ser445Arg	NS		0	0	1	<0.01
10	1344delG	FS	NS		0	0	1	<0.01
11	1411G → A	Glu471Lys	NS		0	0	3	<0.01
11	1513delATCAC	FS	NS		0	0	1	<0.01
12	1622T → C	Leu541Pro	0.001	Yes	0	0	11	0.01
13	1804C → T	Arg602Trp	NS		0	0	3	<0.01
13	1805G → A	Arg602Gln	NS		0	0	1	<0.01
13	1819G → T	Gly607Trp	NS		0	0	1	<0.01
13	1823T → A	Phe608Ile	NS		0	0	1	<0.01
13	1927G → A	Val643Met	NS		0	0	1	<0.01
14	1989G → T	Trp663Stop	NS		0	0	1	<0.01
14	2005delAT	FS	NS		0	0	3	<0.01
14	2041C → T	Arg681Stop	NS		0	0	2	<0.01
14	2147C → T	Thr716Met	NS		0	0	1	<0.01
15	2291G → A	Cys764Tyr	NS		0	0	1	<0.01
15	2294G → A	Ser765Asn	NS		0	0	1	<0.01
15	2300T → A	Val767Asp	NS		0	0	2	<0.01
16	2385del16bp	FS	NS		0	0	1	<0.01
16	2453G → A	Gly818Glu	NS		0	0	1	<0.01
16	2461T → A	Trp821Arg	NS		0	0	1	<0.01
16	2546T → C	Val849Ala	NS		0	0	4	<0.01
16	2552G → A	Gly851Asp	NS		0	0	1	<0.01
16	2560G → A	Ala854Thr	NS		0	0	1	<0.01
17	2588G → C	Gly863Ala	0.0006	No	2	2	28	0.02
17	2617T → C	Phe873Leu	NS		0	0	1	<0.01
18	2690C → T	Thr897Ile	NS		0	0	1	<0.01
18	2701A → G	Thr901Ala	NS		0	1	0	<0.01
18	2703A → G	Thr901Arg	NS		0	0	2	<0.01
19	2828G → A	Arg943Gln	NS		20	13	37	0.05
19	2883delC	FS	NS		0	0	1	<0.01
20	2894A → G	Asn965Ser	NS		0	0	3	<0.01
19	2912C → A	Thr971Asn	NS		0	0	1	<0.01
19	2915C → A	Thr972Asn	NS		0	0	1	<0.01
20	2920T → C	Ser974Pro	NS		0	0	1	<0.01
20	2966T → C	Val989Ala	NS		0	0	2	<0.01
20	2977del8bp	FS	NS		0	0	1	<0.01
20	3041T → G	Leu1014Arg	NS		0	0	1	<0.01
21	3055A → G	Thr1019Ala	NS		0	0	1	<0.01
21	3064G → A	Glu1022Lys	NS		0	0	1	<0.01
21	3091A → G	Lys1031Glu	NS		0	0	1	<0.01
21	3113G → T	Ala1038Val	0.001	Yes	1	0	17	0.01
22	3205insAA	FS	NS		0	0	1	<0.01
22	3261G → A	Glu1087Lys	NS		0	0	2	<0.01
22	3322C → T	Arg1108Cys	0.04	Yes	0	0	6	<0.01
22	3323G → A	Arg1108His	NS		0	0	1	<0.01
23	3364G → A	Glu1122Lys	NS		0	0	1	<0.01

(*continues*)

Table 1 (*continued*). Thirty-Three Truncated and 98 Amino Acid–Changing Variants in the *ABCA4* Gene

Exon	Nucleotide Change	Effect	(A)	(B)	AMD (*n* = 182)	Control (*n* = 96)	STGD (*n* = 374)	Allele Prevalence
23	3386G → T	Arg1129Leu	NS		0	0	3	<0.01
24	3531C → A	Cys1158Stop	NS		0	0	1	<0.01
25	3749T → C	Leu1250Pro	NS		0	0	1	<0.01
26	3835delGATTCT	FS	NS		0	0	1	<0.01
27	3940C → A	Pro1314Thr	NS		0	1	0	<0.01
28	4139C → T	Pro1380Leu	0.001	Yes	0	0	10	0.01
28	4222T → C	Trp1408Arg	NS		0	0	2	<0.01
28	4223G → T	Trp1408Leu	NS		0	0	2	<0.01
28	4234C → T	Gln1412stop	NS		0	0	1	<0.01
29	4297G → A	Val1433Ile	NS		1	0	0	<0.01
29	4319T → C	Phe1440Ser	NS		0	0	1	<0.01
30	4353 − 1g → t	Splice site	NS		0	0	1	<0.01
30	4457C → T	Pro1486Leu	NS		0	0	1	<0.01
30	4462T → C	Cys1488Arg	NS		0	0	3	<0.01
30	4463G → T	Cys1488Phe	NS		0	0	2	<0.01
30	4469G → A	Cys1490Tyr	NS		0	0	3	<0.01
30	4531insC	FS	NS		0	0	2	<0.01
32	4538A → G	Gln1513Arg	NS		0	0	1	<0.01
30	4539 + 1g → t	Splice site	NS		0	0	1	<0.01
31	4574T → C	Leu1525Pro	NS		0	0	1	<0.01
33	4733delGTTT	FS	NS		0	0	1	<0.01
35	4859delATAACAinsTCC T	FS	NS		0	0	1	<0.01
36	4909G → A	Ala1637Thr	NS		0	0	1	<0.01
35	4918C → T	Arg1640Trp	NS		0	0	1	<0.01
35	4919G → A	Arg1640Gln	NS		0	0	1	<0.01
35	4954T → G	Tyr1652Asp	NS		0	0	1	<0.01
36	5077G → A	Val1693Ile	NS		0	0	1	<0.01
36	5186T → C	Leu1729Pro	NS		0	0	2	<0.01
36	5206T → C	Ser1736Pro	NS		0	0	1	<0.01
36	5212del11bp	FS	NS		0	0	1	<0.01
37	5225delTGGTGGTGGGC	FS	NS		0	0	1	<0.01
37	5278del9bp	del LPA 1760	NS		0	0	1	<0.01
37	5288delG	FS	NS		0	0	1	<0.01
38	5395A → G	Asn1799Asp	NS		0	0	1	<0.01
38	5451T → G	Asp1817Glu	NS		1	0	4	<0.01
39	5584 + 5g → a	Splice site	0.02	Yes	0	0	6	<0.01
40	5603A → T	Asn1868Ile	0.0006	No	20	7	79	0.08
40	5651T → A	Val1884GLu	NS		0	0	1	<0.01
40	5657G → A	Gly1886Glu	NS		0	0	1	<0.01
40	5687T → A	Val1896Asp	NS		0	0	1	<0.01
40	5693G → A	Arg1898His	NS		0	0	1	<0.01
40	5714 + 5g → a	Splice site	NS		0	0	1	<0.01
42	5843CA → TG	Pro1948Leu	NS		11	7	28	0.04
42	5882G → A	Gly1961Glu	<0.0001	Yes	1	0	43	0.03
43	5908C → T	Leu1970Phe	NS		1	0	1	<0.01
43	5917delG	FS	NS		0	0	1	<0.01
44	6079C → T	Leu2027Phe	0.01	Yes	0	0	9	0.01
44	6088C → T	Arg2030Stop	NS		0	0	2	<0.01
44	6089G → A	Arg2030Gln	NS		0	0	1	<0.01
44	6112A → T	Arg2038Trp	NS		0	0	1	<0.01
45	6148A → C	Val2050Leu	NS		1	0	0	<0.01
46	6212A → T	Tyr2071Phe	NS		0	0	1	<0.01
45	6229C → T	Arg2077Trp	NS		0	0	2	<0.01
46	6320G → A	Arg2107His	0.01	Yes	0	0	10	0.01
46	6383A → G	His2128Arg	NS		0	0	1	<0.01
47	6446G → T	Arg2149Leu	NS		0	0	1	<0.01
47	6449G → A	Cys2150Tyr	NS		0	0	5	<0.01
48	6529G → A	Asp2177Asn	NS		2	0	0	<0.01
48	6686T → C	Leu2229Pro	NS		0	0	1	<0.01
48	6707delTCACACAG	FS	NS		0	0	1	<0.01
48	6729 + 1g → a	Splice site	NS		0	0	1	<0.01
49	6764G → T	Ser2255Ile	0.009	No	16	4	54	0.06
49	6788G → T	Arg2263Leu	NS		0	0	1	<0.01

(A) The probability under the null hypothesis of similar prevalence of each variant in Stargardt (STGD) compared with non-STGD alleles (two-tailed Fisher's exact test); (B) compatability of the variant existing in a ratio of 100:1 in STGD to control alleles, calculated using the binomial distribution. Allele prevalence refers to the whole study cohort. NS, nonsignificant.

TABLE 2. Fifty-Eight Intronic Variants in the *ABCA4* Gene

Exon	Nucleotide Change	(A)	(B)	AMD (n = 182)	Control (n = 96)	STGD (n = 374)	Allele Prevalence
3	161 − 19G → A	NS		0	0	1	<0.01
3	161 − 45T → G	NS		0	0	1	<0.01
3	302 + 20C → T	NS		0	0	2	<0.01
3	302 + 26G → A	NS		1	1	9	0.01
7	769 − 32T → C	NS		3	0	10	0.01
8	859 − 11C → T	NS		0	0	1	<0.01
9	1100 − 14T → C	NS		0	0	2	<0.01
10	1240 − 14T → C	NS		2	4	6	0.01
10	1356 + 5DELG	NS		18	3	37	0.04
10	1356 + 6g → c	NS		0	0	1	<0.01
12	1760 + 22G → T	NS		0	0	1	<0.01
13	1761 − 37G → A	NS		0	0	1	<0.01
13	1761 − 50G → A	NS		0	0	1	<0.01
13	1761 − 54G → A	NS		3	9	41	0.04
16	2383 − 10C → G	NS		0	0	2	<0.01
17	2588 − 12c → g	NS		1	0	0	<0.01
17	2653 + 60G → C	NS		6	0	2	0.01
18	2654 − 21A → T	NS		0	0	1	<0.01
18	2654 − 48G → C	NS		4	0	9	0.01
19	2744 − 56G → A	NS		0	0	1	<0.01
20	3050 + 61g → a	NS		0	0	1	<0.01
21	3051 − 14T → A	NS		0	0	3	<0.01
21	3190 + 83A → T	0.02	Yes	0	0	6	<0.01
24	3523 − 12C → T	NS		0	1	0	<0.01
24	3607 + 32G → A	NS		0	0	1	<0.01
25	3608 − 16t → a	NS		0	0	1	<0.01
28	4129 − 35A → T	NS		1	0	0	<0.01
28	4253 + 4C → T	NS		0	0	1	<0.01
29	4254 − 38G → A	0.003	Yes	0	0	10	0.01
29	4254 − 47T → C	NS		3	0	1	<0.01
29	4352 + 32A → G	NS		0	0	4	<0.01
30	4539 + 21delg	NS		1	0	0	<0.01
30	4539 + 35G → C	NS		0	0	1	<0.01
30	4539 + 3G → A	NS		0	0	2	<0.01
30	4539 + 40C → T	NS		0	0	1	<0.01
33	4773 + 48C → T	NS		1	0	2	<0.01
36	5196 + 20G → A	NS		0	0	1	<0.01
37	5312 + 45G → T	NS		0	0	1	<0.01
39	5461 − 10T → C	0.002	Yes	0	1	18	0.01
39	5461 − 51INSG	NS		36	34	85	0.12
40	5585 − 70T → C	<0.0001	No	20	8	79	0.08
41	5715 − 24A → C	NS		3	1	8	0.01
41	5835 + 37A → C	NS		0	0	1	<0.01
42	5836 − 11G → A	NS		52	28	115	0.15
42	5836 − 24g → a	NS		1	0	0	<0.01
42	5836 − 3G → A	NS		0	0	1	<0.01
42	5836 − 43C → A	NS		52	28	117	0.15
42	5898 + 22C → A	NS		2	0	0	<0.01
44	6006 − 16G → A	NS		3	0	4	0.01
44	6006 − 17G → A	NS		1	1	0	<0.01
45	6282 + 7g → a	NS		1	0	0	<0.01
45	6282 + 7G → A	NS		0	4	8	0.01
48	6729 + 21C → T	0.001	Yes	0	0	13	0.01
49	6730 − 27C → G	NS		0	0	2	<0.01
49	6730 − 3T → C	0.03	No	16	4	52	0.06
49	6816 + 28C → G	NS		2	0	4	<0.01
50	6817 − 85C → T	NS		0	0	1	<0.01
50	6823 + 26C → A	NS		1	0	0	<0.01

(A) and (B) are described in Table 1. STGD, Stargart disease; NS, nonsignificant.

heterozygous changes. This is far higher, even after accounting for gene size, than the *VMD2* gene in non-Best subjects (0.14 sites, 0.8×10^{-4} sites/nucleotide) or *EFEMP1* gene in non–Malattia Leventinese–Doyne subjects (0.003 sites, 0.03×10^{-4} sites/nucleotide) who were studied using the same strategy for detecting variants and the same control population.

No protein-truncating variations were detected in any AMD or control alleles. When comparing patients with AMD and control subjects, we detected missense variants in 80 (22%) of 364 AMD alleles compared with 37 (19%) of 192 control alleles. If the three most common missense variants (Asn1868Ile, Ser2255Ile, and Arg943Gln) were not included, missense variants were detected in 34 (9.3%) of 364 AMD alleles and 13 (6.7%) of 192 control alleles. Amino acid substitutions that would be expected to alter the charge, size, or polarity of the ABCA4 protein were found in 59 (16%) of 364

Table 3. Twenty-Four Synonymous Codon Variants in the *ABCA4* Gene

Exon	Nucleotide Change	Effect	(A)	(B)	AMD (n = 182)	Control (n = 96)	STGD (n = 374)	Allele Prevalence
3	234T → C	Asn78Asn	NS		0	0	1	<0.01
8	873G → A	Pro291Pro	NS		0	0	1	<0.01
8	981C → T	Pro327Pro	NS		0	0	1	<0.01
10	1248A → C	Ser416Ser	NS		1	0	0	<0.01
10	1269C → T	His423His	NS		2	0	2	<0.01
12	1623A → G	Leu541Leu	NS		0	0	1	<0.01
12	1653G → A	Val551Val	NS		0	0	1	<0.01
14	2127G → A	Ser709Ser	NS		0	0	1	<0.01
19	2877C → T	Thr959Thr	NS		1	1	0	<0.01
20	2964C → T	Leu988Leu	NS		0	0	4	<0.01
25	3696T → G	Leu1232Leu	NS		0	0	1	<0.01
28	4203C → A	Pro1401Pro	0.0001	No	6	9	56	0.05
30	4506C → T	Cys1502Cys	NS		0	0	2	<0.01
33	4746C → T	Asp1582Asp	NS		0	0	1	<0.01
40	5682G → C	Leu1894Leu	0.002	No	62	30	176	0.21
41	5814A → G	Leu1938Leu	NS		20	7	50	0.06
42	5844A → G	Pro1948Pro	NS		41	22	92	0.12
42	5868C → T	Asp1956Asp	NS		0	0	1	<0.01
44	6069C → T	Ile2023Ile	0.04	No	17	12	62	0.07
45	6249C → T	Ile2083Ile	NS		22	14	55	0.07
45	6255C → T	Leu2085Leu	NS		0	1	1	<0.01
46	6285T → C	Asp2095Asp	<0.0001	No	2	1	30	0.03
46	6345G → A	Val2114Val	NS		0	0	1	<0.01
49	6732G → A	Val2244Val	NS		0	0	1	<0.01

(A) and (B) are described in Table 1. STGD, Stargardt disease; NS, nonsignificant.

alleles compared with 26 (13.5%) of 192 control alleles. None of these comparisons was statistically significant at the 5% level (two-tailed Fisher's exact test). Other classes of variation were distributed similarly between AMD and control alleles (synonymous codon changes: 174/364 versus 97/192, intronic polymorphisms: 234/364 versus 127/192; non-significant Mann–Whitney test). Furthermore, no single *ABCA4* variant was found at a significantly higher prevalence in AMD than in control alleles, even before correcting for multiple measurements. In contrast, Stargardt alleles were significantly enriched for all classes of variants. The median number of variants per patient was significantly higher in the Stargardt group than in the others when missense variants, protein-truncating variations, synonymous codon variants, and intronic variations were considered (Kruskal–Wallis $P < 0.001$ all groups).

Variants Predicted to Cause Premature Protein Termination or Aberrant Splicing

Thirty-three variants would be expected to cause premature termination of translation or aberrant splicing of mRNA. These variants were found in 45 instances of 45 alleles of patients with Stargardt disease. Thus, assuming they are disease causing, they represent 45 (6%) of 748 of Stargardt alleles. No patients with Stargardt disease were found to be homozygous for these variations, although this is not significantly different from the expected proportion of homozygotes of 0.4% (assuming that these alleles are distributed randomly among all alleles in patients with Stargardt disease). These premature protein termination variants included 7 nonsense codon changes, 17 frame-shifting deletions, and 3 frame-shifting insertions. Six splice-site variants were found in 14 patients with Stargardt disease. No truncating variants were found in the subjects without Stargardt. One frame-shifting single nucleotide deletion existed in *trans* with an in-frame deletion of nine bases in one patient with Stargardt disease. However, in our study, no patients with a Stargardt phenotype had two variants predicted to cause premature protein termination or aberrant splicing.

Amino Acid Substitutions and Their Significance

Overall, 97 different amino acid substitutions were detected in this study (Table 1). Of these, 68 were nonconservative as defined by a predicted alteration of the size, charge, or polarity of the ABCA4 protein. However, the nonconservative subset was not more strongly associated with Stargardt disease than the conservative subset, largely because three of the nonconservative changes exhibited a prevalence in non-Stargardt alleles of more than 4% (Asn1868Ile, Arg943Gln, and Ser2255Ile).

The missense variants were distributed evenly throughout the *ABCA4* gene. There were no evident hot spots for nucleotide substitution. Twenty-eight of the 97 amino acid substitutions occurred in the NBDs of the ABCA4 protein (15 in NBD1 and 13 in NBD2). Of these, eight substitutions altered residues that were conserved in the eukaryotic protein HisP (*Salmonella typhimurium*) as well as both ABCA4 NBDs. A further six amino acid substitutions occurred at sites identical in both ABCA4 NBDs. Of the amino-acids substitutions that disrupt conserved sequences within the NBDs, all were found only on Stargardt-associated chromosomes.

Despite the large degree of polymorphism in humans, the amino acid conservation between human and murine and human and bovine *ABCA4* genes is high (89% and 88% respectively, excluding the carboxyl terminal 37 residues that are present only in the murine gene). The majority of amino acid substitutions occurred at sites conserved among all three species. Only 13 of 97 amino acid substitutions involved codons that were not conserved among all three species. Two of these 13 were the common variants Ser2255Ile and Pro1948Leu.

Comparison of Allele Distribution with Different Models of Disease Mechanism

The distribution of *ABCA4* variants in patients and control subjects was evaluated in the context of two models (see the Methods section). In model A, all DCVs were assumed to be equally pathogenic and were also assumed to be in Hardy–Weinberg equilibrium. The prevalence of each variant was

tested against the two hypotheses H_1 and H_2, depending on their prevalence in Stargardt and non-Stargardt alleles. Eight different amino acid substitutions had distributions compatible with H_2 but not H_1 (suggesting that they are highly penetrant Stargardt disease alleles). These included three nonconservative changes, Gly1961Glu, Arg1108Cys, and Arg212Cys, and five other changes that were conservative by our criteria, Leu541Pro, Ala1038Val, Pro1380Leu, Leu2027Phe, and Arg2107His. Collectively, these eight substitutions occurred 117 times in 97 patients with Stargardt disease but only twice in two patients without Stargardt disease. Some of these changes were found together on the same disease chromosomes (described later).

Eleven amino acid substitutions exhibited an equal prevalence in both Stargardt and non-Stargardt alleles and were thus incompatible with H_2, essentially excluding these variants from being disease causing under the assumptions of model A. These included two common polymorphisms Arg943Gln and Pro1948Leu that occurred on more than 3% of all alleles. This group also included rare changes that occurred on only one or two non-Stargardt alleles. The remaining 75 missense changes had such a low overall prevalence that their distribution did not support either H_1 or H_2 in a statistically significant manner.

Three missense variants, Gly863Ala, Asn1868Ile, and Ser2255Ile, were significantly enriched among patients with Stargardt disease but not to the extent that would be expected if they were fully penetrant Stargardt alleles. Thus, the distribution of these three variants is incompatible with both hypotheses H_1 and H_2 and is difficult to explain in the context of model A. However, the distribution of these alleles could be explained if they have limited pathogenicity, as proposed in model B. That is, such alleles would not be expected to cause disease if paired with themselves or each other, but could cause disease if paired with another allele of higher pathogenicity (e.g., a null allele). The variant Gly863Ala occurred on 28 Stargardt alleles in 28 patients with Stargardt disease. It was also found on two alleles of patients with AMD and two alleles of control subjects. This variant was never homozygous. Of the 30 patients with Stargardt disease, it was paired with a null allele in three patients, but this proportion of null alleles (3/30) is not significantly greater than the proportion in the whole Stargardt allele pool (45/748; $P = 0.4$, Fisher's exact test). The Asn1868Ile variant was found in 71 patients with Stargardt disease and was homozygous in 8 of these. Of these 71 patients, 9 also exhibited null alleles, the proportion of which was not significantly greater than that in the Stargardt allele pool (Fisher two-tailed $P = 0.9$). The variant was also heterozygous in 20 patients with AMD and 7 control subjects. The Ser2255Ile was found in 48 patients with Stargardt disease and was homozygous in 6 of these. Of these 48 patients, 6 also harbored null alleles, the proportion being similar to that found in the whole Stargardt allele pool (Fisher two-tailed $P = 0.12$). It also occurred in 15 patients with AMD (1 homozygous for the change) and 4 control subjects. The two variants Asn1868Ile and Ser2255Ile occurred together in seven patients with Stargardt disease, one being homozygous for Ser2255Ile and heterozygous for Asn1868Ile. They also occurred together in one patient with AMD who was homozygous for Ser2255Ile and heterozygous for Asn1868Ile.

Linkage Disequilibrium between Variants

Many of the polymorphisms in the *ABCA4* gene were found in linkage disequilibrium with each other. For instance, Asn1868Ile was always associated with CTC at codon 1894 (Leu→Leu). Similarly, the Gly1961Glu variant was always associated with either of two rare alleles at codon 1948 (CCG, CTG), which in turn were always found on chromosomes bearing the common intron 41 polymorphism. Furthermore, the missense changes Gly863Ala and Arg943Gln were commonly found together. These two changes occurred together in 27 patients with and 4 without Stargardt disease. However, linkage disequilibrium was not complete, because the rarer Gly863Ala change occurred by itself in three patients with Stargardt disease. Although it was not always possible to confirm phase in all the double heterozygotes, in the patients in which this was possible the two changes were always in *cis*. Similarly, Leu541Pro occurred in 10 patients with Stargardt disease who harbored the variant Ala1038Val, and, when phase could be determined, these variations were found in *cis*. One patient with Stargardt disease bore Leu541Pro without the Ala1038Val variant, and seven harbored the Ala1038Val change without Leu541Pro. Ala1038Val without Leu541Pro was also found in one patient with AMD.

Detection Rate of Stargardt Variants and Sensitivity of SSCP

One factor that has seriously complicated the analysis of *ABCA4* sequence variation in the context of Stargardt disease is the frequent occurrence of more than two variants in the same patient. (i.e., more than one on the same allele). Considering only those variants predicted to cause protein truncation and missense changes compatible with disease association, we found 1 patient with five such variants, 11 with three, and a further 6 with two (Leu541Pro and Ala1038Val) that were on the same chromosome. Thus, despite detecting a large number of *ABCA4* variants in patients with Stargardt disease, only a relatively small fraction of these appeared likely to be HPRDCVs.

Of the 748 potential Stargardt alleles, 47 harbored protein-truncating variants, and 2 had in-frame deletions. One hundred twenty-six missense changes occurred only in patients with Stargardt disease but at too low a frequency to permit developing any statistically significant support for their role in the pathophysiology of the disease. Accounting for those instances in which two changes were found in *cis*, as well as instances in which three or more changes were found in the same patient, and including those 126 rare missense variants observed only in patients with Stargardt disease, we calculated an upper limit of 263 (35%) of 748 alleles for the proportion of Stargardt alleles carrying changes compatible with being HPRDCVs.

To determine whether this relatively low rate of allele detection could be explained on the basis of the sensitivity of the SSCP assay, the entire *ABCA4* coding region was sequenced in seven patients with Stargardt disease randomly selected from among the group already known to harbor one HPRDCV (and therefore perhaps more likely to harbor another). With this approach, we detected 5 instances of 4 variants in addition to the 41 instances of variation that had been identified with SSCP alone. This gave an estimate of SSCP sensitivity relative to direct sequencing of 89%: $[(41 \times 100)/46]$ (95% confidence interval: 79%–96%; see the Methods section). To determine whether the low DCV detection rate was due to a proportion of patients having macular disease of another cause, 180 Stargardt probands who were examined by at least one of the authors and judged to have changes that were characteristic of the disease were selected from the entire cohort. Using identical criteria, the number of mutant alleles detected in this pool of 360 potential alleles was 154 (43%). In an attempt to identify a specific *ABCA4* allele that was significantly enriched in patients with Stargardt disease (and particularly in those without a detectable DCV) we tested all combinations of pairs and triplets of the 20 most frequent variants in the *ABCA4* gene in the entire study population. No pair or triplet of variants was significantly enriched in patients with Stargardt disease overall,

nor in the subset of patients with Stargardt disease without detectable sequence variations.

Discussion

The identification of any gene involved in the pathogenesis of a human disease raises several exciting possibilities: sensitive and specific molecular diagnosis, accurate prognosis early in the course of the disease, improved understanding of the disease mechanism to the point that conventional therapies can be used more effectively, and true molecular therapy (e.g., gene replacement). However, the full realization of these possibilities requires a thorough characterization of the relationship between genotype and phenotype from at least two perspectives: (1) identification of the range of disease phenotypes that are caused by variations in the gene and (2) differentiation of alleles that truly contribute to a phenotype from those that do not.

In this study, we set out to examine a large cohort of patients with macular disease (and a group of ethnically similar control subjects) for the seemingly straightforward purpose of identifying disease-causing alleles in the *ABCA4* gene. Shortly after the study began, it became apparent that the attainment of this goal would not be nearly as straightforward as it had been for other eye disease genes we had studied. First of all, the gene exhibited a degree of sequence variation that was unprecedented in our experience. Overall, we detected more than 2400 instances of sequence variation in 654 individuals (an average of 3.8 variants per individual for all study participants, including control subjects). When we used an objective statistic to quantify this allelic diversity, we found it to be from 9 to 400 times greater than two other macular disease-causing genes that we have recently characterized. Perhaps the most striking feature of this diversity is that it is unusual for an individual in the general population to be homozygous for the consensus *ABCA4* sequence. The majority of control individuals (76%) harbored at least one variant, and one control individual was found to have eight. It is also noteworthy that of the variants observed in this study, 161 have not been previously reported. This is in agreement with three previous studies of *ABCA4* sequence variation that identified 53,[3,33] 14,[34] and 13[35] variants that were unique to the individual study. This fact emphasizes the need to be aware of differences in genetic variation between different ethnic groups when performing epidemiologic studies. For instance, in the study from the United Kingdom[35] 13 of the total 31 reported variants are at this time specific to the U.K. population. Collectively, this study and the others just noted[3,33–35] have identified a total of 298 different *ABCA4* variants. This diversity makes the detection of Stargardt carriers and the molecular confirmation of the Stargardt diagnosis problematic.

It is of interest that intronic and synonymous coding variants were also significantly enriched in the Stargardt cohort compared with the non-Stargardt cohort. There are a number of possible explanations that may alone or together explain this observation. First, a proportion of the intronic and synonymous codon changes may themselves be pathogenic, although the mechanism by which such seemingly innocuous changes may cause disease is presently unknown. This may apply to four intronic variants (3190 + 83a → t, 4254 − 38g → a, 5461 − 10t → c, 6729 + 21c → t) that are enriched in patients with Stargardt disease in a ratio compatible with an HPRDCV; Table 2). Second, a proportion of the seemingly innocuous variants may be in linkage disequilibrium with undetected disease-causing changes and may be subsequently enriched in patients with Stargardt disease. Finally, it is acknowledged that the ascertainment of patients with Stargardt disease is necessarily wider than that for the patients with AMD and control subjects. Therefore, this may represent a greater degree of ethnic diversity, which is presently not possible to accurately quantify and therefore results in a greater number of alleles and thus the nucleotide variation assayed.

An important hypothesis has been put forth[36,37] that the *ABCA4* gene may play a role in the pathogenesis of the late-onset macular disease known as AMD. As discussed more fully elsewhere,[38] we did not find any statistically significant association between heterozygous changes in the *ABCA4* gene and the AMD phenotype in the cohort of patients in the present study nor in a second larger cohort that included patients from Canada, Europe, and Australia.[39]

A major obstacle to inferring the pathogenicity of the sequence variations we observed in the present study was the autosomal recessive inheritance pattern of Stargardt disease. Unlike rare autosomal dominant macular diseases in which HPDDCVs would be expected to be extremely rare in a control population, HPRDCVs would be expected to be present at an easily detectable frequency in the control population.

A final obstacle we encountered as we tried to correlate specific genotypes with the Stargardt phenotype is that the assay we used is incapable of detecting a large number of DCVs in the *ABCA4* gene. That is, regardless of the criteria we used for inferring pathogenicity and despite the extraordinarily large number of sequence variations we observed, we were unable to identify plausible DCVs in a large number of Stargardt alleles. There are several possible explanations for this observation. The first, and most likely, is that a number of DCVs exist outside the regions of the gene that we screened in the study. For example, deletions affecting one or both primer-binding sites would be refractory to detection using any polymerase chain reaction (PCR)–based protocol. A large deletion was recently detected in one patient by using Southern blot analysis,[34] but it remains to be seen whether such deletions are common in patients with Stargardt disease. A very large deletion (of the entire *ABCA4* gene) would be invisible to both Southern- and PCR-based strategies. However, such deletions are unlikely to be present in a large number of patients in this study, because every patient that we found to harbor any *ABCA4* sequence variation was heterozygous for at least one of their variants. A deletion of the entire gene would result in apparent homozygosity (in reality, hemizygosity) for all variants on the remaining allele.

Another possible explanation for our inability to find more HPRDCVs is genetic heterogeneity among the patients who were included in the Stargardt group. This is unlikely for at least two reasons. First, there was an insignificant increase in the fraction of patients with HPRDCVs when the diagnostic criteria were made more stringent. More important, the distribution of the HPRDCVs that we observed was inconsistent with genetic heterogeneity. That is, if genetic heterogeneity was the sole explanation for observing putative HPRDCVs on only 35% of the alleles of the Stargardt cohort, we should have observed two DCVs in 35% of the patients and none in the remainder. Instead, we observed two disease-causing alleles in 18%, one allele in 35%, and neither allele in 47%. That the heterozygotes outnumbered the homozygotes argues strongly against genetic heterogeneity as the principal explanation for the low HPRDCV detection rate.

Yet another possible explanation for the low HPRDCV detection rate is that our SSCP assay may have failed to detect a large number of missense variations that would have been detected if we had used a different screening technique, such as automated DNA sequencing. This is also unlikely for two reasons. First, the sensitivity of SSCP analysis of amplimers that are 190 bp or less in length has been shown to be higher than 90%.[29] In addition, when we completely sequenced the coding

regions of seven patients in whom one HPRDCV had been found (and who collectively harbored 41 instances of *ABCA4* sequence variation), we detected only five additional instances of variation, and only one of these was compatible with an HPRDCV.

A final possible explanation for failing to find the majority of DCVs in our patients with Stargardt disease is that we have used the wrong model to interpret the genotypic data and to assign disease-causing status to the observed alleles. For example, Arg943Gln is present in more than 6% of the normal population. If it were an HPRDCV, it alone would be expected to cause disease in more than 1 in 300 individuals, a rate that is 30 times higher than the prevalence of Stargardt disease in the population. If, on the other hand, Arg943Gln is a low-penetrance recessive DCV that is incapable of causing disease in the homozygous state but is capable of contributing to the disease phenotype when paired with certain rare HPRDCVs, this variant alone would increase our DCV detection rate by approximately 5%.

Thus, the low detection rate of DCVs in this study could be somewhat artifactual if we are incorrectly discounting the pathogenicity of low-penetrance sequence variants on the basis of their relatively high frequency in the normal population. Although such a phenomenon is suggested by the significant enrichment in Stargardt probands of all types of *ABCA4* variants, we were unable to identify any combinations of variants in the Stargardt probands in this study that were consistently associated with disease. However, it should be noted that it was not possible to determine the phase of the variants in a large proportion of our patients with Stargardt disease, and knowledge of phase is critical to such an analysis. It is interesting to note that some common variants in another ABC transporter gene (encoding the cystic fibrosis transmembrane conductance regulator CFTR), are known to affect the expression of other variants in the same gene.[40,41]

When examining the likelihood of the distribution of variants in patients with and those without Stargardt disease, it is of interest that three missense changes were incompatible with both an equal distribution in the groups (H_1) and a 100% association with Stargardt alleles. One likely explanation for this is that these changes are in partial linkage disequilibrium with DCVs that have remained undetected. The missense change Gly863Ala is of particular interest, because it was found in 4% of Stargardt alleles and in 0.7% of non-Stargardt alleles. Although this distribution is skewed, it is significantly different from the 1:100 ratio expected for a recessive allele for a disease affecting 1 in 10,000 of the population. Maugeri et al.[34] recently showed that a patient heterozygous for this variant (and having a splice variant on the other *ABCA4* allele that appeared to eliminate mRNA expression) showed mRNA on reverse transcription–polymerase chain reaction (RT-PCR) of both a deleted glycine amino acid due to mis-splicing and a Gly/Ala amino acid substitution. It is uncertain therefore whether either of these effects is compatible with normal protein function, particularly given the conservative nature of the amino acid substitution. In the study by Maugeri et al.[34] and others,[3,33,35] this variant was labeled a DCV because of its enrichment in Stargardt alleles.

In a gene such as *ABCA4*, the screening of new patients and families in whom Stargardt disease is suspected, or who are thought to carry Stargardt-associated alleles, is problematic. Unlike some other macular dystrophies, such as Sorsby fundus dystrophy and Malattia Leventinese, in which there is a very limited genetic variation, routine genetic testing of patients with Stargardt disease is not really practical at this time. Sequencing of the whole gene in every patient would be extremely expensive, and the resultant data would be difficult to interpret because of the allelic diversity of the *ABCA4* gene. One approach is to collect data from various studies and choose those variants that appear to be most common in patients with Stargardt disease. Each of these common variants can then be assayed individually, either by using a restriction enzyme, if the variant creates a novel site, or by using an allele-specific PCR assay, if this is technically possible.

One note of caution is that a Stargardt-associated allele in one population, may no longer be in complete linkage disequilibrium with the disease in a different population, as we have recently shown for the most common change in white patients with Stargardt disease, Gly1961Glu, which appears to be quite common in normal individuals from Somalia.[39] Further in vitro studies of mutated *ABCA4* proteins, such as those reported by Sun et al.[1] might help determine the dysfunction of, and thus the potential pathogenicity of, specific *ABCA4* alleles. Clearly, the *ABCA4* gene represents a challenge to all those interested in using molecular techniques to supplement clinical diagnostic methods and to increase the accuracy of genetic counseling for patients with retinal degenerations.

Acknowledgments

The authors thank Luan Streb and Heidi Haines for their excellent technical assistance; and Jean Bennett, Kevin Brady, Ronald Carr, Arlene Drack, Gilbert Grand, Alessandro Iannacone, Howard Schatz, Mitchel Wolf, Lawrence Yanuzzi, Michael York, and Ehud Zamir for sharing their patients with us for this study.

References

1. Sun H, Molda RS, Nathans J. Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific atp-binding cassette transporter responsible for Stargardt disease. *J Biol Chem.* 1999;274:8269–8281.
2. Weng J, Mata NL, Azarian SM, Tzekov RT, Birch DG, Travis GH. Insights into the function of rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in ABCR knockout mice. *Cell.* 1999;98:13–23.
3. Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet.* 1997;15:236–246.
4. Sun H, Nathans J. Stargardt's ABCR is localized to the disc membrane of retinal rod outer segments. *Nat Genet.* 1997;17:15–16.
5. Molday LL, Rabin AR, Molday RS. ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat Genet.* 2000;25:257–258.
6. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other atp-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1982;1:945–951.
7. Hyde SC, Emsley P, Hartshorn MJ, et al. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature.* 1990;346:362–365.
8. Allikmets R, Gerrard B, Hutchinson A, Dean M. Characterization of the human abc superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. *Hum Mol Genet*. 1996;5:1649–1655.
9. Rommens JM, Iannuzzi MC, Kerem B, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science.* 1989;245:1059–1065.
10. Strautnieks SS, Bull LN, Knisely AS, et al. A gene encoding a liver-specific abc transporter is mutated in progressive familial intrahepatic cholestasis. *Nat Genet.* 1998;20:233–238.
11. Martinez–Mir A, Paloma E, Allikmets R, et al. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet.* 1998;18:11–12.
12. Cremers FP, Van De Pol DJ, Van Driel M, et al. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet.* 1998;7:355–362.

13. Blacharski PA. Fundus flavimaculatus. In: Newsome DA, ed. *Retinal Dystrophies and Degenerations.* New York: Raven Press; 1988:135–159.
14. Fishman GA, Farber M, Patel BS, Derlacki DJ. Visual acuity loss in patients with Stargardt's macular dystrophy. *Ophthalmology.* 1987;94:809–814.
15. Kaplan J, Gerber S, Larget-Piet D, et al. A gene for Stargardt's disease (fundus flavimaculatus) maps to the short arm of chromosome 1. *Nat Genet.* 1993;5:308–311.
16. Zhang K, Kniazeva M, Han, M, et al. A 5-bp deletion in elovl4 is associated with two related forms of autosomal dominant macular dystrophy. *Nat Genet.* 2001;27:89–93.
17. Stone EM, Nichols BE, Kimura AE, Weingeist TA, Drack A, Sheffield VC. Clinical features of a Stargardt-like dominant progressive macular dystrophy with genetic linkage to chromosome 6q. *Arch Ophthalmol.* 1994;112:765–772.
18. Edwards AO, Miedziak A, Vrabec T, et al. Autosomal dominant Stargardt-like macular dystrophy, I: clinical characterization, longitudinal follow-up, and evidence for a common ancestry in families linked to chromosome 6q14. *Am J Ophthalmol.* 1999;127:426–435.
19. Lois N, Holder GE, Fitzke FW, Plant C, Bird AC. Intrafamilial variation of phenotype in Stargardt macular dystrophy-fundus flavimaculatus. *Invest Ophthalmol Vis Sci.* 1999;40:2668–2675.
20. Mantyjarvi M, Tuppurainen K. Stargardt's disease: family studies. *Doc Ophthalmol.* 1992;79:79–89.
21. Aaberg TM. Stargardt's disease and fundus flavimaculatus: evaluation of morphologic progression and intrafamilial co-existence. *Trans Am Ophthalmol Soc.* 1986;84:453–487.
22. Isashiki Y, Ohba N. Fundus flavimaculatus: polymorphic retinal change in siblings. *Br J Ophthalmol.* 1985;69:522–524.
23. Rozet JM, Gerber S, Ghazi I, et al. Mutations of the retinal specific atp binding transporter gene (abcr) in a single family segregating both autosomal recessive retinitis pigmentosa RP19 and Stargardt disease: evidence of clinical heterogeneity at this locus. *J Med Genet.* 1999;36:447–451.
24. Lotery AJ, Munier FL, Fishman GA, et al. Allelic variation in the *VMD2* gene in Best disease and age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2000;41:1291–1296.
25. Stone EM, Lotery AJ, Munier FL, et al. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. *Nat Genet.* 1999;22:199–202.
26. Bird AC, Bressler NM, Bressler SB, et al. An international classification and grading system for age- related maculopathy and age-related macular degeneration: the International ARM Epidemiological Study Group. *Surv Ophthalmol.* 1995;39:367–374.
27. Azarian, SM, Megarity CF, Weng J, Horvath DH, Travis GH. The human photoreceptor rim protein gene (ABCR): genomic structure and primer set information for mutation analysis. *Hum Genet.* 1998;102:699–705.
28. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA.* 1989;86:2766–2770.
29. Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics.* 1993;16:325–332.
30. Richard I, Beckmann JS. How neutral are synonymous codon mutations? (Letter). *Nat Genet.* 1995;10:259.
31. Alward WL, Fingert JH, Coote MA, et al. Clinical features associated with mutations in the chromosome 1 open- angle glaucoma gene (GLC1)A. *N Engl J Med.* 1998;338:1022–1027.
32. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA.* 1979;76:5269–5273.
33. Lewis RA, Shroyer NF, Singh N, et al. Genotype/phenotype analysis of a photoreceptor-specific atp-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet.* 1999;64:422–434.
34. Maugeri A, Van Driel MA, Van De Pol DJ, et al. The 2588g→c mutation in the ABCR gene is a mild frequent founder mutation in the western European population and allows the classification of abcr mutations in patients with Stargardt disease. *Am J Hum Genet.* 1999;64:1024–1035.
35. Papaioannou M, Ocaka L, Bessant D, et al. An analysis of *ABCR* mutations in British patients with recessive retinal dystrophies. *Invest Ophthalmol Vis Sci.* 2000;41:16–19.
36. Allikmets R, Shroyer NF, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science.* 1997;277:1805–1807.
37. Allikmets R. Further evidence for an association of ABCR alleles with age-related macular degeneration. *Am J Hum Genet.* 2000;67:487–491.
38. Stone EM, Webster AR, Vandenburgh K, et al. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. *Nat Genet.* 1998;20:328–329.
39. Guymer RH, Héon E, Lotery AJ, et al. Variation of codons and 2177 of the Stargardt disease gene (ABCA4) is not associated with age-related macular degeneration. *Arch Ophthalmol.* In press.
40. Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nat Genet.* 1993;3:151–156.
41. Cuppens H, Lin W, Jaspers M, et al. Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes: the polymorphic (tg)m locus explains the partial penetrance of the t5 polymorphism as a disease mutation. *J Clin Invest.* 1998;101:487–496.