Title
Analysis of the ABCA4 genomic locus in Stargardt disease

Journal
Human Molecular Genetics, 23(25)

ISSN
0964-6906

Authors
Zernant, Jana
Xie, Yajing
Ayuso, Carmen
et al.

Publication Date
2014-12-20

DOI
10.1093/hmg/ddu396

Peer reviewed

Human Molecular Genetics, 2014, Vol. 23, No. 25 **6797–6806**
doi:10.1093/hmg/ddu396
Advance Access published on July 31, 2014

Analysis of the *ABCA4* genomic locus in Stargardt disease

Jana Zernant[1], Yajing (Angela) Xie[1], Carmen Ayuso[3,4], Rosa Riveiro-Alvarez[3,4], Miguel-Angel Lopez-Martinez[3,4], Francesca Simonelli[5], Francesco Testa[5], Michael B. Gorin[6,7], Samuel P. Strom[6,7,8], Mette Bertelsen[9], Thomas Rosenberg[9], Philip M. Boone[10], Bo Yuan[10], Radha Ayyagari[11], Peter L. Nagy[2], Stephen H. Tsang[1,2], Peter Gouras[1], Frederick T. Collison[12], James R. Lupski[10], Gerald A. Fishman[12] and Rando Allikmets[1,2,*]

[1]Department of Ophthalmology and [2]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA, [3]Department of Genetics, Instituto de Investigacion Sanitaria-University Hospital Fundacion Jimenez Diaz, UAM (IIS-FJD), Madrid, Spain, [4]Centro de Investigacion Biomedica en Red (CIBER) de Enfermedades Raras, ISCIII, Madrid, Spain, [5]Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy, [6]Department of Ophthalmology, [7]Department of Human Genetics, Jules Stein Eye Institute and [8]Department of Pathology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, [9]Kennedy Center Eye Clinic, Glostrup Hospital, Glostrup, Denmark, [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, [11]Department of Ophthalmology, University of California San Diego, La Jolla, CA, USA and [12]The Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse for People Who are Blind or Visually Impaired, Chicago, IL, USA

Received May 29, 2014; Revised May 29, 2014; Accepted July 29, 2014

Autosomal recessive Stargardt disease (STGD1, MIM 248200) is caused by mutations in the *ABCA4* gene. Complete sequencing of *ABCA4* in STGD patients identifies compound heterozygous or homozygous disease-associated alleles in 65–70% of patients and only one mutation in 15–20% of patients. This study was designed to find the missing disease-causing *ABCA4* variation by a combination of next-generation sequencing (NGS), array-Comparative Genome Hybridization (aCGH) screening, familial segregation and *in silico* analyses. The entire 140 kb *ABCA4* genomic locus was sequenced in 114 STGD patients with one known *ABCA4* exonic mutation revealing, on average, 200 intronic variants per sample. Filtering of these data resulted in 141 candidates for new mutations. Two variants were detected in four samples, two in three samples, and 20 variants in two samples, the remaining 117 new variants were detected only once. Multimodal analysis suggested 12 new likely pathogenic intronic *ABCA4* variants, some of which were specific to (isolated) ethnic groups. No copy number variation (large deletions and insertions) was detected in any patient suggesting that it is a very rare event in the *ABCA4* locus. Many variants were excluded since they were not conserved in non-human primates, were frequent in African populations and, therefore, represented ancestral, and not disease-associated, variants. The sequence variability in the *ABCA4* locus is extensive and the non-coding sequences do not harbor frequent mutations in STGD patients of European-American descent. Defining disease-associated alleles in the *ABCA4* locus requires exceptionally well characterized large cohorts and extensive analyses by a combination of various approaches.

INTRODUCTION

Mutations in the *ABCA4* gene are responsible for a wide variety of retinal dystrophy phenotypes from autosomal recessive Stargardt disease (STGD1) (1) to cone–rod dystrophy (CRD) (2,3) and, in some advanced cases, retinitis pigmentosa (RP) (2,4,5). While CRD and RP phenotypes are also caused by mutations in many other genes, *ABCA4* is the only recognized gene

*To whom correspondence should be addressed. Email: rla22@columbia.edu

responsible for STGD1 (MIM 248200), a predominantly juvenile-onset macular dystrophy frequently associated with early-onset central visual impairment, progressive bilateral atrophy of the foveal retinal pigment epithelium, and the presence of yellowish flecks, defined as lipofuscin deposits, around the macula and/or in the central and near-peripheral areas of the retina.

Over 800 disease-associated *ABCA4* variants have been already identified (6) and the most frequent of these have been described in only ~10% of STGD1 patients (7). Several studies have identified frequent 'ethnic group-specific' *ABCA4* alleles, such as the p.G863A/G863del founder mutation in Northern European patients (8), the p.[L541P;A1038V] complex allele in patients of mostly German origin (3,9), the p.R1129L founder mutation in Spain (10), the p.N965S variant in the Danish population (11) and the p.A1773V variant in Mexico (12).

Complete sequencing of the *ABCA4* coding and adjacent intronic sequences in patients with STGD1 routinely discovers ~80% of mutations with the fraction of patients harboring the expected two disease-associated alleles at 65–70%, with one mutation ~15–20%, and with no mutations in the remaining ~15% (13). These fractions depend on many variables, most importantly the quality of the clinical diagnosis and the ethnic composition of the cohort. Most of the cases with no detected *ABCA4* mutations likely represent phenocopies (13); i.e. in those patients mutations in other gene(s) cause a STGD1-like phenotype. However, based on the known carrier frequency of pathogenic *ABCA4* variants, in most cases with one *ABCA4* allele the second allele is expected to reside in the *ABCA4* locus. It can present as a copy number variant (CNV, large deletion or insertion of one exon or more) which eludes detection by PCR-based sequencing techniques, a synonymous variant in the coding region, or a (deep) intronic variant, which may affect splicing or a regulatory region, such as a promoter or an enhancer (14). Very few of these have been identified (13,15,16).

This study was designed to find the missing *ABCA4* mutations by a combination of next-generation sequencing, array-comparative genomic hybridization (aCGH) arrays, *in silico* and RNA analyses, and segregation analyses in families.

RESULTS

Discovery of new disease-associated variants by next-generation sequencing

Sequencing of the entire *ABCA4* genomic locus, at an average depth of coverage of 100×, in 130 patients with *ABCA4*-associated disease harboring one previously known *ABCA4* disease-associated allele, and 6 patients with no known *ABCA4* mutations, resulted in detecting 1745 different variants. Eighty-three of these were previously known disease-associated or benign variants from coding regions and pathogenic splice site variants. Six hundred and ninety-five (695) variants were also detected in 1000 Genomes Project or Exome Sequencing Project, with no statistically significant differences in allele frequencies between the general population and the patient cohort, unless the variants were on the same allele (haplotype) with the frequent known *ABCA4* coding mutation, p.G1961E. Five hundred and twenty-six (526) variants were incorrectly called deletions or insertions from single nucleotide repeat areas (homopolymers) that have proven to be difficult for the NGS approach. We also experienced a relatively high A>C/C>A/T>G/G>T false-positive calling rate with Illumina sequencing. The number of false positives can be reduced by more stringent criteria for variant calling; however, this may also exclude some real variants. After the filtering and verification steps 141 new intronic *ABCA4* variants remained in 114 patients. In 22 patients with one previously known *ABCA4* mutation, the second pathogenic *ABCA4* allele was also found in the coding sequence or adjacent splice sites. In 6/22 cases this was due to re-evaluation of several variants which had been classified as benign, e.g. p.G991R and p.A1773V. The remaining 16 cases represented false-negative results, probably due to technical reasons in the first, sequencing, step of the ABCA4 coding regions.

Of the 141 new possible candidates for disease-associated variants, two variants, c.4539+2064C>T and c.5461-1389C>A, were detected together (one the same chromosome) as a complex allele in four patients of Spanish or Italian descent (Table 1). The c.5461-1389C>A variant is in an evolutionarily less conserved area, the c.4539+2064C>T variant is adjacent to the recently reported c.4539+2028C>T and c.4539+2001G>A variants from a conserved area (14). According to predictive programs, none of these variants have any effect on splicing, whether on existing cryptic splice sites or on creating new sites. The c.4539+2064C>T and c.5461-1389C>A haplotype segregated with the disease in all three STGD1 families from Spain (Fig. 1A–C), and were absent in 100 matched Spanish control samples, making these variants very likely candidates for intronic *ABCA4* mutations.

Two variants, c.5196+1056A>G and c.6006-609T>A, were detected in 3/114 unrelated patients each and were absent in 368 matched control samples. The c.5196+1056A>G variant segregated with the disease in two families; i.e. it was on a different chromosome than the proband's other *ABCA4* mutation (Fig. 1D). This variant is also predicted to strengthen a cryptic splice donor (Supplementary Material, Fig. S1A). The variant was found, in addition, in 1/119 patients from our replication cohort of STGD1 patients with one known *ABCA4* mutation and has been recently reported as a disease-associated allele (14). The aggregate evidence suggests that the c.5196+1056A>G variant is a rare, deep intronic disease-associated allele.

The c.6006-609T>A variant, which does not have a predicted effect on splicing, was detected in 6/119 additional unrelated samples from our replication cohort of STGD1 patients with one known *ABCA4* mutation. Two of these samples were from the Columbia University patient cohort and four from European cohorts including two from Denmark. The variant was not present in 368 European-American control samples, but was detected in 2/182 Danish control samples. While the frequency of this variant is 10× elevated in STGD1 patients as compared with all controls (1.9 versus 0.18%) and 3.5× if compared with Danish controls (0.55%), it is premature to unequivocally call the variant as associated with the disease.

Three variants, c.570+1798A>G, c.2161-8G>A, and c.859–9T>C were each detected in 2/114 different patients in our primary *ABCA4* locus screening cohort, but absent in 368 control samples and also in 119 additional STGD1 samples from the replication cohort (Table 1). The c.570+1798A>G

Table 1. Analysis of the new intronic *ABCA4* variants which were either detected twice or more in the cohort of 114 STGD1 patients, and/or have a predicted effect on splicing

Position on chr1	Variant	Effect on splicing (combined Alamut prediction)	C score	Primary *ABCA4* locus cohort (114)	Validation (replication) cohort (119)	Segregation with disease	Controls[a]	Conservation in primates	Disease association
94 580 645	c.67-2023T>G	Cryptic donor strongly activated	1.888	1	0		0/368	No	Probably not
94 566 773	c.570+1798A>G	New donor site	8.955	2	0		0/368	Yes	Yes
94 563 992	c.768+358C>T	No effect	9.324	2	0		0/368	Yes	Possibly yes
94 550 775	c.769-1778T>C	No effect	4.085	2	0		3/368	Yes	No
94 546 283	c.859-9T>C	Weakens the acceptor by ~14%	4.495	2	0	Yes	0/368, 0/120[b]	Yes	Yes
94 526 934	c.1938-619A>G	Cryptic donor strongly activated	2.211	1	0	Yes	0/368	Yes	Yes
94 525 509	c.2160+584A>G	New donor site	12.53	1	0		0/368	Yes	Yes
94 522 386	c.2161-8G>A	Weakens the acceptor by ~50%	13.25	2	0		0/368	Yes	Yes
94 514 521	c.2654-8T>G	Weakens the acceptor by ~35%	11.3	1	0		0/368	Yes	Yes
94 511 126	c.2919-826T>A	Cryptic donor strongly activated	1.513	1	0		0/368	No	Probably not
94 509 799	c.3050+370C>T	New donor site	1.072	1	0		0/368	Yes	Yes
94 495 923	c.4352+61G>A	Cryptic donor strongly activated	6.914	1	0		0/368	No	Probably not
94 493 272	c.4539+1729G>T	New donor site	1.586	1	0		0/368	Yes	Yes
94 493 000	c.4539+2001G>A[c]	No effect	4.653	2	0		0/368	Yes	Possibly yes
94 492 973	c.4539+2028C>T[c]	No effect	3.531	2	2		NS	Yes	Possibly yes
94 492 937	c.4539+2064C>T	No effect; the most frequent new variant	3.598	4	0	Yes	0/100[d]	Yes	Yes
94 484 082	c.5196+1056A>G[c]	Cryptic donor strongly activated	2.025	3	1	Yes	0/368	Yes	Yes
94 484 001	c.5196+1137G>A[c]	Cryptic acceptor strongly activated	2.681	1	9		0/368, 8/200[e]	No	No
94 478 330	c.5461-1389C>A[f]	No effect; the most frequent new variant	4.805	4	0	Yes	0/368, 0/100[d]	No	No
94 473 856	c.5836-3C>A	Weakens the acceptor by ~26%	11.84	1	0		NS	No	Yes
94 471 747	c.6006-609T>A	No effect, second most frequent variant	7.22	3	6		0/368, 2/180[g]	Yes	Possibly yes
94 468 019	c.6148-471C>T	Cryptic donor strongly activated	8.47	1	0		0/368	Yes	Yes
94 462 292	c.6730-541T>C	Cryptic acceptor strongly activated	5.682	1	0	No	0/368	Yes	No

[a]Control cohort of 368 samples of European ancestry was screened if not indicated differently.
[b]Cohort of 120 control samples of Asian-Indian origin.
[c]Variant is previously reported by Braun *et al*. (2013).
[d]Cohort of 100 controls from Spain.
[e]Cohort of 200 African-American general population controls.
[f]Variant is on the same allele with c.4539+2064C>T.
[g]Cohort of 180 controls from Denmark. NS, not screened.

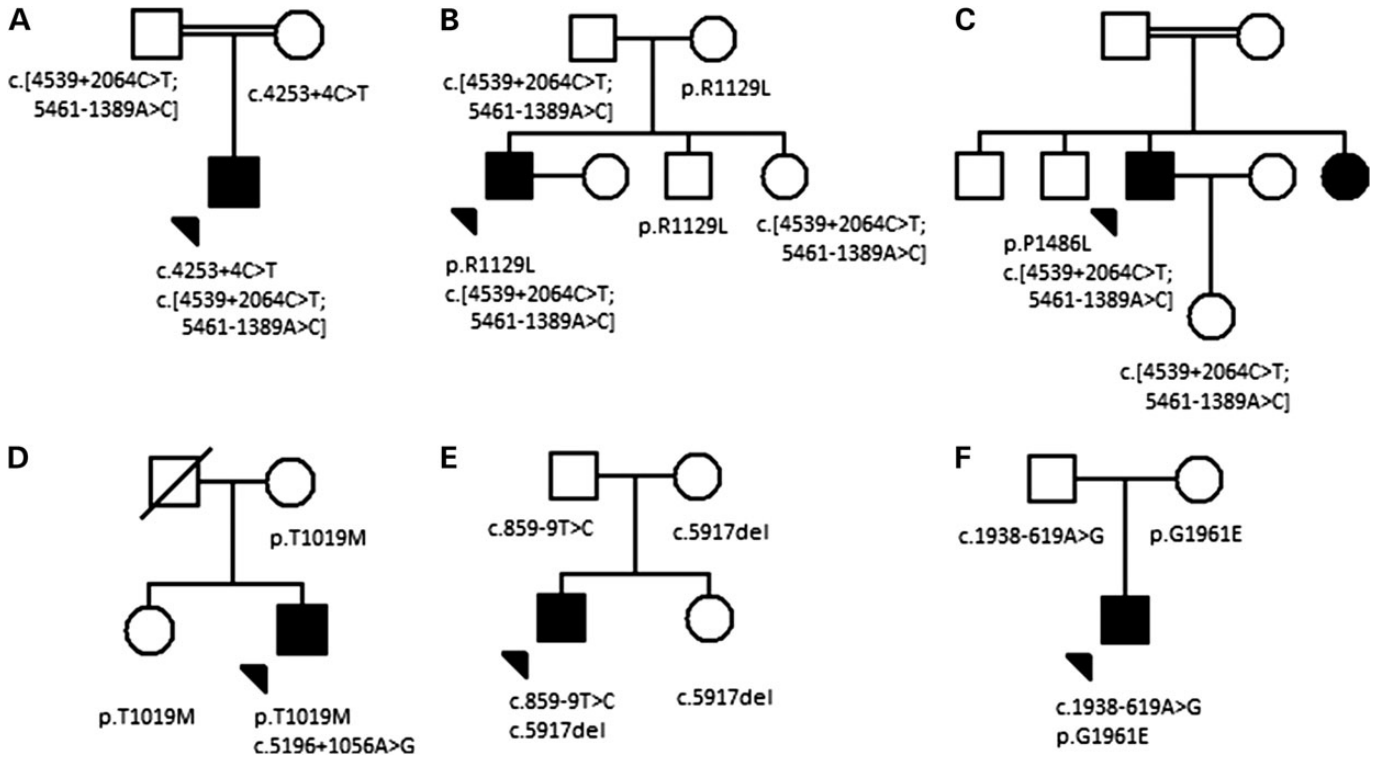

Figure 1. Pedigrees segregating the new *ABCA4* intronic variants with STGD1.

variant creates a very strong new donor splice site according to all predictive software (Supplementary Material, Fig. S1B). The c.2161-8G>A variant weakens the existing acceptor by 50%, and the c.859-9T>C change by 14% (Supplementary Material, Fig. S1C and D). The c.2161-8G>A and c.859-9T>C variants are adjacent to *ABCA4* exon sequences and can be detected also by sequencing of *ABCA4* coding regions. However, neither of these variants has been detected in the Exome Sequencing Project currently containing 4300 individuals of European-American descent and 2203 individuals of African-American descent. In addition to the 2/114 patients from the *ABCA4* locus screening cohort, c.859-9T>C has been detected twice, in homozygosity, in STGD1 patients with no other known *ABCA4* mutations, and in three more STGD1 patients in heterozygous state with one known *ABCA4* mutation. Segregation analysis was possible in one family, and confirmed that the c.859-9T>C was not on the same chromosome with the proband's other mutation (Fig. 1E). All evidence points to these three variants as being intronic disease-associated *ABCA4* variants; the c.859-9T>C variant is discussed in detail below.

Seventeen more new variants were detected twice in 114 patients from the locus screening cohort. Five of these variants were detected together in the same two patients both of whom also carried the most likely benign p.V931M variant, so these variants were eliminated from the pool of possible mutation candidates. Three more variants were on the same chromosome with the previously known *ABCA4* exon variants, (with p.R212C, p.T1253M, and p.[L541P;A1038V]), respectively. Other variants were either in the same patients who already carried a stronger intronic mutation candidate (5), not conserved in non-human primates (NHPs) (1), or were found in controls with similar frequency (1). None of them had a predicted effect on splicing. Among these variants are also the recently reported by Braun *et al*. c.4539+2028C>T and c.4539+2001G>A variants. Both variants were found in 2/114 samples in our primary cohort. The c.4539+2028C>T change was also detected in 2/119 patients of the replication cohort, a frequency significantly lower than reported by Braun *et al*. (14).

The remaining 117 new *ABCA4* intronic variants were only detected once each in 62 different patients with one previously known *ABCA4* mutation. Twelve of these variants were predicted to have an effect on splicing (Table 1). The c.5836-3C>A and c.2654-8T>G weaken the existing splice acceptor sites on average by 25 and 35%, respectively (Supplementary Material, Fig. S1E and F). Neither of these two variants has been detected by Exome Sequencing Project, nor in our entire STGD1 patient cohort (780 patients) where all *ABCA4* coding regions and adjacent splice sites have been sequenced.

The variants c.67-2023T>G, c.1938-619A>G, c.3050+370C>T, c.4352+61G>A and c.2919-826T>A, all strengthen the existing cryptic splice donors sites (Supplementary Material, Fig. S1G and H). None of these variants were found in 368 control samples, nor in the replication cohort of 119 additional STGD1 patients with one *ABCA4* disease-associated allele. The c.1938-619A>G variant segregated with the disease in one family (Fig. 1F). The c.67-2023T>G and c.2919-826T>A variants are not conserved in non-human primates and, therefore, not likely disease-causing in humans. At the same time, these two patients do not carry any other possible *ABCA4* mutant alleles. The variants c.2160+584A>G and c.4539+1729G>T are predicted to create new strong splice donors (Supplementary Material, Fig. S1I and J). The variants are

absent in 368 control samples and in the replication cohort of 119 STGD1 patients (Table 1). Since these positions are highly conserved among species we suggest that the c.2160+584A>G and c.4539+1729G>T variants are very good candidates for intronic *ABCA4* mutations.

The variants c.5196+1137G>A, c.6148-471C>T and c.6730-541T>C have a predicted effect of strengthening the existing cryptic splice acceptors (Supplementary Material, Fig. S1K–M), and were each all detected in one patient in the primary locus screening cohort. The c.6730-541T>C variant was on the same chromosome with the probands other known *ABCA4* mutation, and can therefore be excluded from possible new mutations list. The c.6148-471C>T and c.6730-541T>C variants were not detected in 368 control samples, nor in 119 additional STGD1 samples. The recently reported (14) c.5196+1137G>A variant was additionally found in nine STGD1 samples with one previously known *ABCA4* variant (Table 1). Three of these patients were from Denmark, three were of African-American origin, and three of unknown ethnicity. Screening matched control cohorts revealed no carriers for this allele in 180 Danish control samples, but it was detected in 8/200 (MAF = 2%) general population controls of African-American descent, strongly suggesting non-pathogenicity. Additional support for this variant not being pathogenic comes from evolutionary analysis (see below). The human major allele c.5196+1137G is the minor allele in *Macaca mulatta* and the suggested mutant allele A is the major allele in macaques.

Analysis of regulatory sequences

To assess the effect of the new *ABCA4* intronic variants on putative regulatory regions we compared their location against the chromosome coordinates of the DNaseI hypersensitivity and transcription factor binding regions from the ENCODE project. Since regulatory regions, in particular promoters, tend to be DNase sensitive, variants that fall within such regions may affect regulatory potential. Also, variants that are located in transcription factor binding sites may potentially have an effect on protein binding. The combined DNaseI hypersensitivity data were derived from 125 cell types, and the dataset for transcription factor binding regions involved 161 transcription factors in 91 cell types. Unfortunately the datasets did not contain eye-specific cell types or transcription factors. The defined regions were assigned normalized scores in the range of 0–1000, with higher scores indicating stronger signal strength. Twenty-four of the 141 new *ABCA4* intronic variants are located in regions with both DNaseI hypersensitivity and transcription factor binding scores of various strength (Supplementary Material, Table S1). Family members were available for segregation analyses in three cases. Variants c.570+1499C>A and c.67-3166C>T were on the same chromosome with the probands' other mutation, while c.571-1707C>T and c.5018+289T>C were both detected in one patient and on the different chromosome than this patient's other mutation, suggesting possible pathogenicity. Twenty-one variants fall within regions with only DNaseI hypersensitivity score, and 12 variants within regions of transcription factor binding consensus sequences.

The new *ABCA4* intronic variants were also subjected to the Combined Annotation Dependent Depletion (CADD) algorithm (http://cadd.gs.washington.edu/score) (17). The CADD algorithm combines a diverse array of annotations into one metric (C score) for each variant, ranking a variant relative to all possible substitutions of the human genome (17). C score correlates with allelic diversity, pathogenicity and experimentally measured regulatory effects. A C score of >10 indicates that the variant is among the top 10% of most deleterious substitutions in the human genome. C scores for the new *ABCA4* intronic variants range from 0.004 to 15.14 (Supplementary Material, Table S1), with nine variants resulting in a C score greater than 10. Four of those variants were already classified as possibly disease-associated due to their strong predicted effect on splicing (Table 1). The other five variants with higher C scores included c.4539+1971T>C, c.67-3779T>G, c.303-906A>G, c.859-41T>C and c.1357-221T>A (Supplementary Material, Table S1).

In summary, we found a very strong or a probable intronic mutation candidate in 27/114 (23.7%) patients with one existing definite *ABCA4* mutation (Table 1). No immediately plausible candidates for intronic mutations were found in 36/114 (31.6%) patients. The remaining 51 (44.7%) patients possessed one or more new intronic variants that were only detected once, had no effect on splicing according to prediction programs and, therefore, are very difficult to confirm or refute as disease-associated alleles with the available methods and the impossibility of obtaining the patient RNA. However, it is highly likely that a fraction of these are associated with STGD1.

Analysis of previously reported variants

Recently several *ABCA4* intronic variants were suggested to account for a substantial fraction of pathogenic *ABCA4* alleles (14). Since we have directly sequenced the *ABCA4* gene and flanking intronic sequences in >780 STGD1 patients and the entire *ABCA4* genomic locus in 114 STGD patients with one

Table 2. Frequency of the variants described in Braun *et al*. (2013), and in the current study.

Variant # (Braun *et al*.)	Location on chr1	Variant	Braun *et al*. primary cohort (*n*=28)	CU locus cohort (*n* = 114)	CU exon seq cohort (*n* = 780)
V1	94 484 001	c.5196+1137G>A	4	1	–
V2	94 483 922	c.5196+1216C>A	1	0	–
V3	94 484 082	c.5196+1056A>G	1	3	–
V4	94 493 000	c.4539+2001G>A	1	2	–
V5	94 492 973	c.4539+2028C>T	3	2	–
V6	94 466 602	c.6342G>A/p.V2114V	1	0	1
V7	94 487 399	c.4773+3A>G	3	0	0

mutation, we compared the Braun *et al*. data to ours (Table 2). The silent exonic p.V2114V variant, which we had described in our earlier study as a possibly disease-associated mutation (13), is very rare; it was detected in 1 out of 780 STGD1 patients (Table 2). The 'near-exonic' c.4773+3A>G variant was not detected in any of our 780 patients. We were also unable to detect one of the remaining five 'deep intronic' variants (c.5196+1216C>A) in any patient with one mutation and the other four were seen, in total, in 9/114 (7.9%) STGD1 patients, a statistically significant difference from the Braun *et al*. data 10/28 (17.9%; $P = 0.001$).

We then analyzed the evolutionary and/or ethnic origin of the variants focusing first on the most frequent variant in the Braun *et al*. study, c.5196+1137G>A. Evolutionary conservation of a nucleotide is one of the most important criteria for determining the pathogenicity of a variant, especially in a highly conserved gene such as *ABCA4*. The ABCA4 protein performs a very specialized function in the visual cycle; therefore it is exceptionally conserved in mammals and in all vertebrates with visual cycle. For example, the mouse and human ABCA4 proteins are 88% identical, allowing the human protein to perform the transport function in mouse (18). The conservation extends beyond coding sequences and includes splice sites; in fact, the *ABCA4* gene has the same structure, consisting of exactly the same 50 exons in non-human primates, such as *Pan troglodytes* and *M. mulatta*. The evolutionary conservation extends, in some regions, deep into the introns; for example, the 200 bp sequences surrounding the c.5196+1137G>A variant are 96% identical between human and macaque. Most importantly, *M. mulatta* has the adenosine nucleotide in human position c.5196+1137 (Fig. 2), as a major allele with guanosine as a minor allele; i.e. a situation exactly the reverse to that observed in humans, suggesting that this is an ancestral variant and not a disease-causing mutation. To further prove this assumption, we screened 200 unrelated individuals from the general population of African-American descent and identified eight heterozygotes, resulting in the allele frequency of ~2%. The cohort of STGD1 patients at Columbia University contains 46 African-American patients of whom one carried the c.5196+1137G>A variant, resulting in the allele frequency of ~1%, comparable to that in the general population. No other disease-associated *ABCA4* variant was identified in this patient. In addition, we detected the c.5196+1137G>A variant only once in 114 STGD1 patients (allele frequency 0.4%) of European descent in the Columbia cohort (Tables 1 and 2). However, it was found in 3 out of 24 STGD1 patients (allele frequency 6.25%) in a STGD1 cohort from Denmark (Table 1). Interestingly, all three patients derived from the same region in Denmark, suggesting either admixture, or that this variant may be more frequent also in some (isolated) ethnic groups other than of African descent since the variant was not detected in 180 unaffected Danish individuals from Copenhagen. Finally, RNA analysis from macaque homozygous for the A allele at the human position c.5196+1137 clearly showed no effect on splicing (Fig. 2), which eliminates the c.5196+1137G>A variant as possibly pathogenic in STGD1 patients.

The three remaining deep intronic variants described by Braun *et al*. (14), c.5196+1056A>G, c.4539+2028C>T and c.4539+2001G>A are discussed in the previous section. All of these are very rare and are conserved in NHPs. The c.5196+1056A>G variant, which affects splicing, is likely a

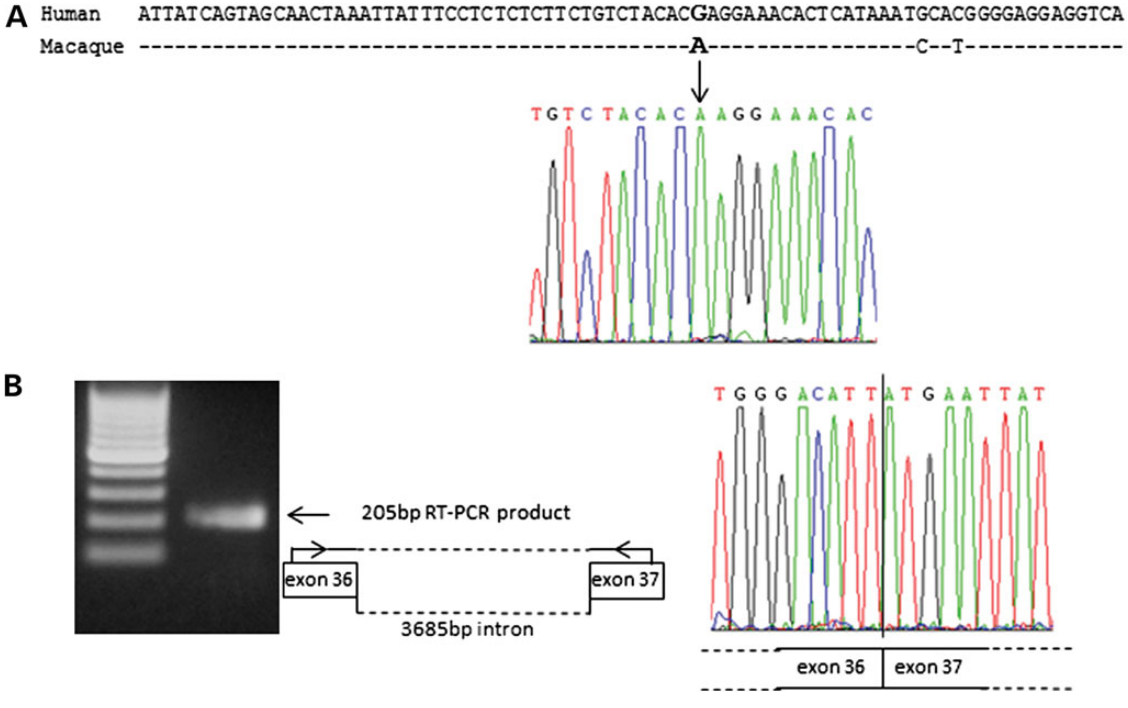

Figure 2. Analysis of the c.5196+1137G>A variant in *M. mulatta*. (**A**) Alignment of the *Homo sapiens* chromosome1: 94 484 046–94 483 967, GRCh37.p13 Primary Assembly, with *M. mulatta* respective sequences. *ABCA4* intronic position c.5196 + 1137G is marked with large font in bold. Differences in macaque sequence compared with human are designated with letters. (**B**) Confirmation of correct splicing of *ABCA4* exons 36 and 37 in a c.5196+1137A macaque retina in RT-PCR analysis. No alternate splicing products were detected.

disease-associated mutation. The other two are not predicted to affect splicing and are too rare to investigate by other means, so the pathogenicity of these variants remains unconfirmed.

New frequent intronic variant in STGD1 patients of Asian Indian descent

As described above, we detected the c.859-9T>C variant in 2/114 patients in our primary *ABCA4* locus screening cohort. Since the c.859-9T>C variant weakens the existing splice acceptor only by 14%, we initially did not consider the variant a strong candidate for disease association. However, subsequently the variant was detected in homozygous state in two STGD1 patients who did not harbor any other exonic or intronic *ABCA4* mutations and, heterozygously, in three more STGD1 patients with one known *ABCA4* mutation. Review of the ethnic origin of all these patients determined that they were all of Asian Indian descent originating from either Pakistan, India or Bangladesh. Segregation analysis was possible in one family, and confirmed that the c.859-9T>C segregated with the disease (Fig. 1E); i.e. it was in *trans* configuration with the second *ABCA4* mutation, c.5917del, in the proband. The c.859-9T>C variant is adjacent to *ABCA4* exon sequences and is, therefore, detected also by exome sequencing, or by direct sequencing of *ABCA4* coding regions. However, it has not been detected in the Exome Sequencing Project currently containing 4300 individuals of European-American descent and 2200 individuals of African-American descent. It was also absent in 368 control samples and also in 119 additional STGD1 samples of European-American descent with one known *ABCA4* mutation (Table 1). Further screening in Asian-Indian population did not detect this variant in 120 subjects, both from the general population (50 individuals), or from the patients affected with Leber congenital amaurosis or Leber hereditary optic neuropathy (70 cases). Altogether, the data suggest that the variant is not frequent in the ethnically matched general population and is very frequent in Indian patients with STGD1 (7 alleles in 15 patients—23.3%) suggesting that the c.859-9T>C variant is a frequent disease-associated *ABCA4* allele in patients of Asian-Indian origin.

Analysis of the copy number variants by aCGH arrays

Few lines of evidence have suggested that some disease-associated *ABCA4* alleles can present as CNVs, mostly in the form of large deletions encompassing one or more exons (15). These reports have been rare, so we reasoned that a small fraction of missing *ABCA4* alleles might present as CNVs. In total of 104 STGD1 patients with one known *ABCA4* allele and 5 patients with no *ABCA4* mutations were screened on the custom CGH arrays, 57 of these were also included in the locus sequencing. aCGH data of the total number of 109 samples was analyzed by using Agilent Genomic Workbench (version 7.0), where CNVs were called by using ADM-2 algorithm with threshold of 4.0. No large (>500 bp) CNVs were identified in the *ABCA4* locus, while ultra-small, seemingly true-positive CNV calls were further validated by PCR, which did not confirm any, reflecting the decreased accuracy of aCGH for the ultra-small CNVs Since the array contained several genes in the *CFH* locus with known frequent CNVs and confirmed in a positive control a heterozygous 1030 bp deletion in the *ABCA4* locus, we can exclude any technical issues and conclude that, despite reports of CNVs (mainly large deletions) in 1–2% of STGD1 patients, these events are likely much more rare and the CNVs do not account for a reasonable fraction of missing *ABCA4* alleles in all populations studied.

DISCUSSION

The *ABCA4* gene was described as the causal gene for STGD1 17 years ago (1) and has been subjected to intensive genetic research; however, the complete understanding of genetic causality has yet to be elucidated. Major factors challenging genetic analysis are: (1) the size of the gene, (2) the exceptional genetic heterogeneity, (3) the expression of the gene in only photoreceptors, rods and cones and, consequently and most importantly, (4) the impossibility of obtaining retinal tissue samples *in vivo* from patients for RNA analyses. Until recently, even the analysis of amino acid-changing variants was stymied by the lack of a direct functional assay; these analyses were limited to indirect *in vitro* studies of the transporters' ability to bind and/or hydrolyze ATP (19,20). This problem has been somewhat alleviated by the recent studies of Molday *et al.* (21,22), where direct transport assays, albeit still *in vitro*, have been proposed and successfully used for several *ABCA4* missense alleles. Large-scale use of the assay for hundreds of documented *ABCA4* variants is still time- and cost-prohibitive. However, the functional analyses of variants that do not affect the protein sequence directly, such as splicing defects and variants in regulatory regions likely modulating levels of expression, remain refractory to experimental verification.

Some studies (9,14) have employed the 'illegitimate transcript' or 'minigene' strategies to assess the effect of certain variants on splicing, but these studies have serious limitations (in part stemming from the word 'illegitimate') and the results, while suggestive, cannot be considered unequivocal, as also demonstrated in the current study. With the lack of patient RNA for direct structural and expression studies, the analysis is limited to the assessment of variant frequencies in STGD1 patients and in the matched general populations, to *in silico* suggestions by predictive software programs, and to segregation analyses in families. The latter approach is also seriously hampered by the fact that most variants in *ABCA4* coding and non-coding regions are extremely rare, most often represented in singleton cases and the available families/pedigrees are usually nuclear, i.e. mostly very small, where the segregation analysis has a very limited power.

In the current study of non-coding genetic variation on the largest STGD1 cohort analyzed to date, we were able to (almost) unequivocally assign pathogenicity to only 12/141 variants in intronic sequences of *ABCA4*. Most of the variants which occurred only once in the *ABCA4* locus cannot be called especially when the predictive programs do not suggest any effect on splicing. Moreover, the assessment based on predictive programs is not unequivocal, several studies have suggested that molecularly confirmed predictions range from 70 to 80% (23). One has to take into account the fact that assessment of the effect on splicing addresses just one possible mechanism for non-coding sequences. Effect on these on other regulatory

elements affecting the gene expression, such as transcription factor binding sites, enhancers, promoters, etc., is still very difficult to assess in most cases (24). The best known example of an intronic variant in the *ABCA4* locus is the c.5461-10T>C variant in intron 38, which is the second or third most frequent variant (found in ~7% of STGD1 patients of European descent) after the p.G1961E and p.[L541P;A1038V] mutations (13). The c.5461-10T>C variant always segregates with the disease phenotype in families, is very rare in the general population (<0.001) and is shown not to affect splicing in a minigene approach (9). Since there is no mutation in the *ABCA4* coding sequence on the same chromosome with the c.5461-10T>C variant, the latter has to be a disease-associated mutation, although its functional consequences remain unknown.

A recent study suggested a few deep intronic variants in the *ABCA4* locus which were associated with STGD1 in a large fraction, up to 50%, of patients with one exonic mutation. Our detailed analysis of these data with a multi-faceted approach on a much larger cohort of European-American descent could not confirm these conclusions. Specifically, the reported variants were much, 10–50 times, rarer than originally claimed (Table 2) and some of them were deemed not associated with STGD1 after evolutionary and RNA analyses employing NHPs (Fig. 2). Concerns also include the ethnic origin and relatedness of patients in cohorts described by Braun *et al*. (14), since the statistically significant difference in allele frequencies between two studies can occur, excluding technical issues, for two main reasons: (1) cohorts include related subjects or, (2) difference in the racial or ethnic origin of the cohorts. For example, recently several missense *ABCA4* variants that were originally considered pathogenic since they were very rare in populations of European descent have now been classified non-pathogenic since they are major alleles (i.e. the 'wild-type' variants) in non-human primates (e.g. p.R1300Q) and/or frequent in some racial and ethnic groups, for example, in African-Americans (e.g. p.L1201R, p.R1300Q and p.V643M) (25).

In summary, the analysis of the entire *ABCA4* locus in large cohorts of STGD1 patients revealed the following:

(1) The genetic variability in the non-coding sequences in the *ABCA4* locus, similar to the coding sequences, is exceptionally vast.
(2) There are no frequent pathogenic variants in the non-coding sequences of the *ABCA4* locus; all definitely or likely disease-associated variants are individually rare in the populations of European descent.
(3) Some variants are more frequent in specific racial and ethnic groups.
(4) Copy number variations in the *ABCA4* locus are very rare.

Analysis of the pathogenicity of specifically intronic *ABCA4* variants, which affect splicing and regulatory regions influencing the gene expression, is complicated due to the impossible task of obtaining RNA from photoreceptor cells from affected individuals *in vivo*. Studying iPS cells obtained from individual patients, which are then directed towards differentiating into photoreceptors with the goal of expressing *ABCA4*, could be a plausible approach.

MATERIALS AND METHODS

Patients

STGD1 patients (255) were, after written informed consent, recruited and clinically examined during a 10-year period in different centers in the USA, Italy, Spain and Denmark. Control cohorts included samples from centers in the USA, Spain and Denmark. In total, 255 patients, and 918 control samples were included in the analyses.

Our primary locus sequencing cohort of 136 samples consisted of 49 STGD1 patients from Columbia University, 22 from the University of Illinois at Chicago and the Pangere Center at the Chicago Lighthouse, 22 from UCLA, 25 patients from Italy and 18 from Spain. The second, validation/replication cohort consisted of 119 STGD1 patients from the same centers. Age of onset was defined as the age at which symptoms were first reported. Visual acuity was measured using the Early Treatment Diabetic Retinopathy Study Chart 1 or a Snellen acuity chart. Clinical examination, fundus photography, fundus autofluorescence and spectral domain-optical coherence tomography (SD-OCT) (Heidelberg Spectralis HRA + OCT) were performed using standard acquisition protocols following pupil dilation with Tropicamide 1% and Neosynephrine 2%. All research was carried out with the approval of the Institutional Review Boards at the respective centers and in accordance with the Declaration of Helsinki.

M. mulatta samples

DNA samples from 86 unrelated rhesus macaques (*M. mulatta*) from NHP colonies at the National Institutes of Health and the Primate Facility at the University of Oregon were genotyped for the c.5196+1137G>A variant.

Sequencing

The first set of 48 patients were sequenced using RainDance microdroplet-PCR target enrichment (RainDance Technologies, Billerica, MA) with subsequent sequencing on Roche 454 platform (454 Life Sciences, Branford, CT). We targeted the genomic region chr1:94 453 727–94 595 732 (GRCh37/hg19 Assembly); including the *ABCA4* genomic locus chr1:94 458 394–94 586 705 and 9027 bp of 5′UTR and 4667 of 3′UTR sequences. The design covered 100% of the targeted area via 473 amplicons of 350–500 bp.

The second set of 95 patient samples were analyzed by the Illumina Truseq Custom Amplicon target enrichment strategy followed by sequencing on Illumina MiSeq platform (Illumina, San Diego, CA). The Illumina design targeted the genomic region 94 456 700–94 591 600, including the *ABCA4* locus, and 4895 bp of 5′UTR and 1694 bp of 3′UTR sequences. The region was divided into nine targets, with seven 500 bp and one 1100 bp gap introduced into the repeating elements. The cumulative target via Illumina design involved 130 319 bp, covered by 94% with 421 amplicons of 425 bp. The next-generation sequencing reads were analyzed using the variant discovery software NextGENe (SoftGenetics, State College, PA). The reads were aligned against the targeted region in the reference genome GRCh37/hg19. For a variant to be called, we required the read containing the variant to match 85% to the

aligned position, the variant to be covered by at least 10 reads and the variant to be present in 20% of all reads aligned to the given position. We also used the overall confidence score of 10 of the NextGENe software as a further filter. We used the previously determined *ABCA4* coding variants as controls to set these filtering criteria. On average, ∼200 variants were called per individual patient.

Analysis of the *ABCA4* variants

All variants and their allele frequencies were compared with the 1000 Genomes database (26), and to the Exome Variant Server (EVS) dataset, NHLBI Exome Sequencing Project, Seattle, WA, USA (http://snp.gs.washington.edu/EVS/; accessed November 2013). New variants that were not recorded in these databases were further analyzed by a combination of predictive *in silico* methods and statistical analyses. The possible effect of all new non-coding *ABCA4* variants on splicing was assessed using five different algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) via Alamut software (http://www.interactive-biosoftware.com). In order to assess the regulatory potential of the new *ABCA4* intronic variants we compared their chromosome coordinates against the predicted regulatory regions from two ENCODE datasets: (1) Combined DNaseI hypersensitivity clusters from 125 cell types ('Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE' http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=366969521&c=chr1&g=wgEncodeRegDnaseClusteredV2, filename: wgEncodeRegDnaseClusteredV2.bed.gz, last accessed on 23 August 2012); and (2) ChIP-seq clustered regions for 161 transcription factors in 91 cell types ('Transcription Factor ChIP-seq V4 (161 factors) with Factorbook motifs from ENCODE'http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=366969521&c=chr1&g=wgEncodeRegTfbsClusteredV3, filename: wgEncodeRegTfbsClusteredV3.bed.gz, last accessed on 21 July 2013). Evolutionary conservation of the variants was noted via UCSC Genome Browser (http://genome.ucsc.edu). The Combined Annotation Dependent Depletion (CADD) algorithm (http://cadd.gs.washington.edu/score) was used to estimate combined predicted general deleteriousness of every variant. The variants' segregation with the disease in available families was analyzed by Sanger sequencing, and screening of patient and control cohorts for allele frequencies in various populations was performed using TaqMan Genotyping technology (Life Technologies, Carlsbad, CA). For genotyping for the c.5196+1137G>A variant we used PCR RFLP with forward primer 5′GTGGGCCTAGCTCCTTTTAT3′, reverse primer 5′GGAGACCAACACAAATGACC3′ (Life technologies, Carlsbad, CA), and the DNA restriction endonuclease BssSI (New England Biolabs, Ipswich, MA)

Array-comparative genome hybridization (aCGH)

Custom aCGH arrays (Agilent Technologies, CA), in a 8 × 60K format, was designed with high-density probes tiling the critical genetic loci of *ABCA4*-associated disease were designed to assess for CNVs involving these loci. The *ABCA4* locus and eight other known genes causing macular disease *(ELOVL4, PRPH2, BEST1, RS1, CNGB3*), and major age-related macular degeneration-associated loci *(CFH, ARMS2, C2/CFB*), were considered as 'primary loci' and therefore probed with ultra-high density throughout the entire genomic length of the genes plus the 5′ promoter regions and 3′ downstream regions, as well as at slightly lower density for flanking 5′ and 3′ conserved regions. Eighteen genetic loci *(C3, APOE, CFI, LIPC, SYN3/TIMP3, CETP, COL8A1, BBX, PLD1, SPEF2, ADAM19, VEGFA, FRK, MEPCE, CHMP7/LOXL2, TGFBR1, NPS, PICK1)* were considered as 'secondary loci' and probed with ultra-high density in exons and with lower resolution throughout each gene plus the flanking 5′ promoter regions and 3′ downstream regions. The details of the array design are described in the Supplementary Material, Table S2.

Array-CGH analysis was performed on DNA from 104 individuals diagnosed with STGD1, for each of whom only one mutation in *ABCA4* had been found by sequencing. A DNA sample with a known, previously reported 1030 bp heterozygous deletion of exon 18 in *ABCA4*, was used as positive control (15). Experimental procedures of aCGH were performed as described previously (27) with minor modifications. Agilent Genomic Workbench version 7.0 software (Agilent Technologies, CA) was used for data analysis, and PCR validations were performed for the plausible true-positive CNVs after being filtered against several criteria.

ABCA4 RNA analysis from rhesus macaque retina

Total RNA was isolated from a snap frozen rhesus macaque (*M. mulatta*) retina using AllPrep DNA/RNA Mini Kit (QIAGEN Cat. No 80204) with a fast spin-column procedure. cDNA synthesis was achieved using TaqMan Reverse Transcription Reagents (Life technologies, Carlsbad, CA) in a 30 min incubation at 48°C. The primer pair for further PCR amplification was designed to encompass *ABCA4* exons 36 and 37, with the forward primer in the exon 36 of the *ABCA4* gene, 5′GATTTTCTCCATGTCCTTCG3′ and the reverse primer in the exon 37, 5′CTTTCTTCTGAAACCCGATG3′, resulting in an amplicon of 205 bp, in the case of correct splicing.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors thank Dr John Fingert for the analysis of a subset of Asian-Indian samples and Dr Dwight Stambolian for providing African-American general population samples.

Conflict of Interest statement. None declared.

FUNDING

This work was supported, in part, by grants from the National Eye Institute/NIH EY021163, EY019861, EY021237 and EY019007 (Core Support for Vision Research); Foundation Fighting Blindness (Owings Mills, MD), Harold and Pauline Price Foundation, unrestricted funds from Research to Prevent Blindness (New York, NY) to the Departments of Ophthalmology, Columbia University and UCLA, the Pangere Family

Foundation, Chicago Lighthouse, FIS PI13/00226 & CIBERER from ISCIII, Madrid, Spain, ONCE and Fundaluce (Spain).

REFERENCES

1. Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A. *et al.* (1997) A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.*, **15**, 236–246.
2. Cremers, F.P., van de Pol, D.J., van Driel, M., den Hollander, A.I., van Haren, F.J., Knoers, N.V., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A. *et al.* (1998) Autosomal recessive retinitis pigmentosa and cone–rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum. Mol. Genet.*, **7**, 355–362.
3. Maugeri, A., Klevering, B.J., Rohrschneider, K., Blankenagel, A., Brunner, H.G., Deutman, A.F., Hoyng, C.B. and Cremers, F.P. (2000) Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone–rod dystrophy. *Am. J. Hum. Genet.*, **67**, 960–966.
4. Martinez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Dean, M., Vilageliu, L., Gonzalez-Duarte, R. and Balcells, S. (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat. Genet.*, **18**, 11–12.
5. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N. and Lupski, J.R. (2001) Null missense ABCR (ABCA4) mutations in a family with Stargardt disease and retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **42**, 2757–2761.
6. Allikmets, R. (2007) Tombran-Tink, J. and Barnstable, C.J. (eds), In *Retinal Degenerations: Biology, Diagnostics and Therapeutics*, Humana Press, Totowa, NJ, in press, pp. 105–118.
7. Burke, T.R., Fishman, G.A., Zernant, J., Schubert, C., Tsang, S.H., Smith, R.T., Ayyagari, R., Koenekoop, R.K., Umfress, A., Ciccarelli, M.L. *et al.* (2012) Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. *Invest. Ophthalmol. Vis. Sci.*, **53**, 4458–4467.
8. Maugeri, A., van Driel, M.A., van de Pol, D.J., Klevering, B.J., van Haren, F.J., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A., Pinckers, A.J. *et al.* (1999) 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.*, **64**, 1024–1035.
9. Rivera, A., White, K., Stohr, H., Steiner, K., Hemmrich, N., Grimm, T., Jurklies, B., Lorenz, B., Scholl, H.P., Apfelstedt-Sylla, E. *et al.* (2000) A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am. J. Hum. Genet.*, **67**, 800–813.
10. Valverde, D., Riveiro-Alvarez, R., Bernal, S., Jaakson, K., Baiget, M., Navarro, R. and Ayuso, C. (2006) Microarray-based mutation analysis of the ABCA4 gene in Spanish patients with Stargardt disease: evidence of a prevalent mutated allele. *Mol. Vis.*, **12**, 902–908.
11. Rosenberg, T., Klie, F., Garred, P. and Schwartz, M. (2007) N965S is a common ABCA4 variant in Stargardt-related retinopathies in the Danish population. *Mol. Vis.*, **13**, 1962–1969.
12. Chacon-Camacho, O.F., Granillo-Alvarez, M., Ayala-Ramirez, R. and Zenteno, J.C. (2013) ABCA4 mutational spectrum in Mexican patients with Stargardt disease: identification of 12 novel mutations and evidence of a founder effect for the common p.A1773V mutation. *Exp. Eye Res.*, **109**, 77–82.
13. Zernant, J., Schubert, C., Im, K.M., Burke, T., Brown, C.M., Fishman, G.A., Tsang, S.H., Gouras, P., Dean, M. and Allikmets, R. (2011) Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol. Vis. Sci.*, **52**, 8479–8487.
14. Braun, T.A., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., Deluca, A.P., Fishman, G.A., Lam, B.L., Weleber, R.G. *et al.* (2013) Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum. Mol. Genet.*, **22**, 5136–5145.
15. Yatsenko, A.N., Shroyer, N.F., Lewis, R.A. and Lupski, J.R. (2003) An ABCA4 genomic deletion in patients with Stargardt disease. *Hum. Mutat.*, **21**, 636–644.
16. Stenirri, S., Battistella, S., Fermo, I., Manitto, M.P., Martina, E., Brancato, R., Ferrari, M. and Cremonesi, L. (2006) De novo deletion removes a conserved motif in the C-terminus of ABCA4 and results in cone–rod dystrophy. *Clin. Chem. Lab. Med.*, **44**, 533–537.
17. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
18. Kong, J., Kim, S.R., Binley, K., Pata, I., Doi, K., Mannik, J., Zernant-Rajang, J., Kan, O., Iqball, S., Naylor, S. *et al.* (2008) Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. *Gene Ther.*, **15**, 1311–1320.
19. Sun, H., Smallwood, P.M. and Nathans, J. (2000) Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat. Genet.*, **26**, 242–246.
20. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N., Wensel, T.G. and Lupski, J.R. (2001) Cosegregation and functional analysis of mutant ABCR (ABCA4) alleles in families that manifest both Stargardt disease and age-related macular degeneration. *Hum. Mol. Genet.*, **10**, 2671–2678.
21. Quazi, F. and Molday, R.S. (2013) Differential phospholipid substrates and directional transport by ATP-binding cassette proteins ABCA1, ABCA7, and ABCA4 and disease-causing mutants. *J. Biol. Chem.*, **288**, 34414–34426.
22. Quazi, F. and Molday, R.S. (2014) ATP-binding cassette transporter ABCA4 and chemical isomerization protect photoreceptor cells from the toxic accumulation of excess 11-cis-retinal. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 5024–5029.
23. Liu, Y.H., Li, C.G. and Zhou, S.F. (2009) Prediction of deleterious functional effects of non-synonymous single nucleotide polymorphisms in human nuclear receptor genes using a bioinformatics approach. *Drug Metab. Lett.*, **3**, 242–286.
24. Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.*, **13**, 469–483.
25. Utz, V.M., Chappelow, A.V., Marino, M.J., Beight, C.D., Sturgill-Short, G.M., Pauer, G.J., Crowe, S., Hagstrom, S.A. and Traboulsi, E.I. (2013) Identification of three ABCA4 sequence variations exclusive to African American patients in a cohort of patients with Stargardt disease. *Am J Ophthalmol*, **156**, 1220–1227. e1222.
26. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
27. Gonzaga-Jauregui, C., Zhang, F., Towne, C.F., Batish, S.D. and Lupski, J.R. (2010) *GJB1*/connexin 32 whole gene deletions in patients with X-linked Charcot-Marie-Tooth disease. *Neurogenetics*, **11**, 465–470.