Analysis of the ABCA4 Gene by Next-Generation Sequencing

Jana Zernant,1 Carl Schubert,1 Kate M. Im,2 Tomas Burke,1 Carolyn M. Brown,1 Gerald A. Fishman,3 Stephen H. Tsang,1,4 Peter Gouras,1 Michael Dean,2 and Rando Allikmets1,4

Purpose. To find all possible disease-associated variants in coding sequences of the ABCA4 gene in a large cohort of patients diagnosed with ABCA4-associated diseases.

Methods. One hundred sixty-eight patients who had been clinically diagnosed with Stargardt disease, cone-rod dystrophy, and other ABCA4-associated phenotypes were pre-screened for mutations in ABCA4 with the ABCA4 microarray, resulting in finding 1 of 2 expected mutations in 111 patients and 0 of 2 mutations in 57 patients. The next-generation sequencing (NGS) strategy was applied to these patients to sequence the entire coding region and the splice sites of the ABCA4 gene. Identified new variants were confirmed or rejected by Sanger sequencing and analyzed for possible pathogenicity by in silico programs and, where possible, by segregation analyses.

Results. Sequencing was successful in 159 of 168 patients and identified the second disease-associated allele in 49 of 105 (~48%) of patients with one previously identified mutation. Among those with no mutations, both disease-associated alleles were detected in 4 of 56 patients, and one mutation was detected in 10 of 56 patients. The authors detected a total of 57 previously unknown, possibly pathogenic, variants: 29 missense, 4 nonsense, 9 small deletions and 15 splice-site-altering variants. Of these, 55 variants were deemed pathogenic by a combination of predictive methods and segregation analyses.

Conclusions. Many mutations in the coding sequences of the ABCA4 gene are still unknown, and many possibly reside in noncoding regions of the ABCA4 locus. Although the ABCA4 gene remains a good first-pass screening option, the NGS platform is a time- and cost-efficient tool for screening large cohorts. (Invest Ophthalmol Vis Sci. 2011;52:8479–8487 DOI: 10.1167/iovs.11-8182)

Mutations in the ABCA4 gene are responsible for a wide variety of retinal dystrophy phenotypes, such as autosomal recessive Stargardt disease (STGD1),1 cone-rod dystrophy (CRD),2,3 and retinitis pigmentosa (RP).2,4,5 STGD1 (Mendelian Inheritance in Man 248200) is a predominantly juvenile-onset macular dystrophy associated with rapid central visual impairment, progressive bilateral atrophy of the foveal retinal pigment epithelium (Fig. 1), and frequent appearance of yellowish flecks, defined as lipofuscin deposits, around the macula or in the central and near-peripheral areas of the retina. In a large fraction of STGD1 patients, a “dark” or “silent” choroid is seen on fluorescein angiography that reflects the accumulation of lipofuscin throughout the retina.

More than 600 disease-associated ABCA4 variants have been identified,6 and the most frequent disease-associated ABCA4 alleles have each been described in only approximately 10% of STGD1 patients. Several studies have identified frequent “ethnic group-specific” ABCA4 alleles, such as the c.2588G>C variant resulting in a dual effect, p.G863A/delG863, as a founder mutation in Northern European patients with STGD1 and a complex allele (two variants on the same chromosome), p.L541P/A1038V, in both STGD1 and CRD patients of German origin (Fig. 2B).6,7 Complex ABCA4 alleles are not uncommon in STGD1. In fact, they are detected in approximately 10% of all STGD patients.8

Allelic heterogeneity has substantially complicated genetic analyses of ABCA4-associated retinal disease. Efforts related to mutation detection and genotyping become especially crucial in genotype/phenotype correlation studies, in which screening of thousands of samples is needed to achieve enough statistical power because multiple rare variants and their combinations must be studied.1 We generated, at the time, a high-throughput and cost-effective screening tool, the ABCA4 genotyping microarray, using solid-phase arrayed primer extension (APEX) technology. The ABCA4 microarray, which has been regularly updated, contains all known disease-associated genetic variants (>500) in the ABCA4 gene. The chip has been used for efficient, systematic screening of patients with ABCA4-associated diseases,9,10 it detects approximately 65% to 75% of all disease-associated alleles. On average, the array screening finds two mutations in approximately 40% of patients diagnosed with “classical” STGD1. Of the rest, one mutation is detected in 40% of patients, whereas no disease-associated allele is found in the ABCA4 coding region in 20% of screened patients.8

Direct Sanger sequencing of the entire ABCA4 coding region detects between 66% and 80% of the alleles10,11 but...
exons and chromosomal segments), have found these in only approximately 1% of all STGD1 patients.\(^6\)

Overall, variation in the \(ABCA4\) locus has emerged as the most prevalent cause of Mendelian retinal disease because approximately 1 of 20 people across all populations carry a potential disease-associated variant in this gene.\(^7,12,15\) Recent advances in developing therapeutic applications for STGD1 in preclinical studies\(^7,18\) suggest that more comprehensive and affordable genetic screening technologies have to be implemented for molecular diagnosis and for selection of patients who would benefit from specific (especially gene-based) therapeutic modalities. Here we describe screening by NGS of a large cohort of \(ABCA4\)-associated patients who had been analyzed by the \(ABCA4\) array and still lacked one or both mutations.

**Patients, Materials, and Methods**

**Patients**

Patients \((n = 168)\) affected with STGD1 \((n = 155)\), CRD \((n = 15)\), and RP \((n = 2)\) were, after providing written consent, recruited and clinically examined over a 10-year period at the Department of Ophthalmology at Columbia University and at the University of Chicago at Illinois. Patients with CRD and RP were included in this study because they all harbored one \(ABCA4\) disease-associated variant after array screening. 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. Clinical examination, fundus photography, fundus autofluorescence (Fig. 1) and spectral domain–optical coherence tomography (SD-OCT) (Spectralis HRA+OCT; Heidelberg Engineering, Heidelberg, Germany) were performed using standard acquisition protocols after pupil dilation with tropicamide 1%. All research was carried out with the approval of the Institutional Review Board of Columbia University and in accordance with the Declaration of Helsinki.

**Array Screening**

Screening with the \(ABCA4\) array had been performed on all patients, followed by direct sequencing to confirm identified changes, as previ-

**Figure 1.** Fundus autofluorescence image of the left eye of patient 3032 harboring \(ABCA4\) variants c.2300T>A (p.V767D) and c.735T>G (p.Y245*). This combination of \(ABCA4\) mutations resulted in early disease onset at 5 years of age. At 16 years of age, the patient was found to have extensive hypoautofluorescence, indicative of atrophy of the retinal pigment epithelium throughout the macula with patchy extension of hypoautofluorescence into the extramacular retina. Note the relative “sparking” of uniform hyperautofluorescence in the peripapillary region.

**Figure 2.** Pedigrees segregating Stargardt disease. (A) An example of a pedigree with pseudodominant inheritance. Father and son are both affected with atSTGD. Mother is a carrier of a frequent c.3322C>T (p.R1108C) mutation. The new c.3522+5delG variant affecting splicing (Table 1) was detected by NGS. (B) An example of a pedigree segregating a frequent complex \(ABCA4\) allele [c.1622T>C; 3113C>T] (p.L541P;A1038V), in which either mutation separately can cause the disease. The c.3543delT frameshift variant was detected by NGS. (C) An example of a pedigree segregating a complex allele in which one variant (c.2894A>G; p.N965S) causes disease and the other, c.4283C>T, p.T1428M, is a benign polymorphism, although it was originally described as a rare mutation in patients of European descent. The new c.3655G>C, p.A1219P variant was detected by NGS.
ously described. Because the array screening had been performed over many years, different versions of the ABCA4 chip had been used, from the least representative (~300 mutations) to the current version (~600 variants).

NGS

All 50 ABCA4 exons were amplified using an amplicon tagging protocol (Access Array; Fluidigm, South San Francisco, CA; http://www.fluidigm.com/products/access-array.html). The integrated fluidic circuit of this system facilitates parallel amplification of 48 unique sequences, in effect preparing 48 sequencing libraries. The primers (Supplementary Table S4, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8182/-/DCSupplemental) for the amplification step were designed in accordance with the guidelines from Fluidigm, using Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm). Universal forward and reverse tag sequences were added to 5’ ends of the designed primer sequences to enable the use of sample-specific barcode primers consisting of 454 sequence tags (all provided by Fluidigm). Every reaction combined both an amplicon tagging and a bar-coding sample tagging (identification) step that enables all 48 amplicons to be multiplexed at the sequencing step, maximizing the usefulness of the sequencer (454 GS-FLX; Fluidigm). Every reaction contained 50 to 100 ng genomic DNA, 5 pmol forward- and reverse-tagged amplification primers, 1 pmol forward and reverse barcode primers, 1 μl loading reagent (Access Array; Fluidigm), reaction buffer with MgCl2 (FastStart High Fidelity; Roche, Indianapolis, IN), enzyme blend (FastStart High Fidelity; Roche), and 1× PCR-grade nucleotide mix.

Each 48 × 48 array generated 48 amplicons in 48 samples (n = 2304); 168 samples were amplified on 4 (5.5) arrays. The amplicons were tagged with the adaptors (454 Titanium; Roche) during the PCR amplification, pooled, and purified, resulting in one amplicon library per array. The resultant libraries (n = 4) were subjected to emulsion PCR and sequenced using the same chemistry (454 Titanium; Roche). Each library was sequenced in one region of a four-region picotiter plate. All the reactions were carried out in accordance with the manufacturer’s protocols. Each region generated between 130,000 and 200,000 high-quality bidirectional reads, resulting in an average 56× coverage per amplicon. The best amplicons were covered with approximately 140 reads; most amplicons yielded, on average, 60 to 80 sequence reads. The total of 112 amplicons with coverage <10 reads (1.46%, 112/7680; 160 samples × 48 amplicons) were resequenced by the Sanger method.

Sequences of barcoded samples were analyzed with the GS Reference Mapper software (http://www.454.com/products-solutions/analysis-tools/gs-reference-mapper.asp), which mapped reads to the reference genome (HG19) and compiled a consensus sequence. All differences compared to the reference sequence are easily viewed with automatic output to separate files for insertions, deletions, and SNPs. High-confidence differences compared to the reference genome are compiled in a separate output file. The output files were converted to .sam files using a Python script kindly provided by Kevin Jacobs and subsequently to sorted .sam files using samtools [PMID: 19505943]. BAM files were then exported into the Integrated Genome Viewer (Broad Institute) for visualization of all identified variants. All identified variants were confirmed by Sanger sequencing.

Sequence Analyses

New missense variants were analyzed with algorithms such as Sorting Intolerant from Tolerant (SIFT; http://sift.jcvi.org/), and Polymorphism Phenotyping (PolyPhen; http://genetics.bwh.harvard.edu/pph/) to predict the impact of variants on the ABCA4 function and, consequently, on disease susceptibility. Variants detected in adjacent to exons intronic sequences were analyzed with splice site prediction programs GeneSplicer (http://www.ccbcb.umd.edu/software/GeneSplicer), and Splice Site Finder (www.genet.sickkids.on.ca/~all/splicefinder). All the prediction programs were accessed with bioinformatics software (Alamut 2.0; http://www.interactive-biosoftware.com). Where available, segregation of the new variants with the disease was analyzed in families (Fig. 2).

RESULTS

Discovery of New Disease-Associated Variants by NGS

The entire ORF (all 50 ABCA4 exons and flanking intronic sequences) was sequenced in 168 ABCA4-associated patients (Supplementary Table S1, http://www.iovs.org/lookup/suppl doi:10.1167/iovs.11-8182/-/DCSupplemental). Of these, 111 had one pathogenic mutation previously identified by the ABCA4 array; in 57 patients array screening had not revealed any mutations, although all of them had been clinically confirmed as affected with ABCA4-associated disease. PCR amplification, barcoding, tagging, and pooling were performed on the Fluidigm 48 × 48 Access Arrays. Screening of the ABCA4 gene is especially amenable for the Access Array system because 50 exons can be amplified as 48 amplicons. Sequencing did not work in eight patients, most likely because of a Fluidigm array error given that all these samples were in one array column. One patient with one identified ABCA4 mutation represented a false negative for the ABCA4 array (i.e., this patient should not have been included in further sequencing analysis because both mutations in this patient should have been identified by the array). Three additional false-negative samples for the ABCA4 array were also detected; however, because no mutations were found, these samples were still valid for sequencing as one-mutation cases. Therefore, nine patients—seven with one mutation, one with no mutations, and one with two mutations—were excluded from the final analyses (Supplementary Table S1, http://www.iovs.org/lookup/suppl doi:10.1167/iovs.11-8182/-/DCSupplemental). New variants were analyzed by a combination of predictive in silico methods, statistical analyses, segregation with the disease in the families, and screening of large patient and control cohorts for population frequencies. All variants were first compared against our most up-to-date ABCA4 mutation database, which includes all published and communicated ABCA4 variants. As predicted, the previously unknown identified variants were rare and not present in public databases such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP) and HapMap (http://hapmap.ncbi.nlm.nih.gov/). Filtering against these databases excluded known frequent variants with frequencies similar to those in the general population.

Of the 103 patients with one mutation, the NGS strategy found the second disease-associated allele in 49 patients (47.6%). Of the 56 patients with no identified mutations, NGS detected both disease-associated alleles in four patients (7.1%) and one mutation in 10 patients (17.9%). We detected a total of 57 previously unknown variants in 64 patients: 29 missense, 4 nonsense, 9 small deletion, and 15 splice-acting variants from the nucleotide changes 5 base pairs upstream or downstream of the exons (Table 1). All but two (55/57) new variants were deemed pathogenic by a combination of predictive in silico methods and segregation analyses in available families (Table 1). A single-nucleotide deletion, c.885delC, was detected in three patients; the splice site mutation c.3522+5delG was found twice in the same family (parent and child each had an STGD1 diagnosis), and the nonsense mutation p.R2149* was detected in two patients. Two splice-acting variants, c.6479+1G>A and c.6479+1G>C, targeted the same position. All the other newly identified mutations were detected only once.

The mutations p.D108V, p.Y245*, p.R2149*, and c.4667+1G>A, which were identified by NGS, were already known and included in the current ABCA4 array but were not at the time these samples were screened. Systematic screening of the
# Table 1. Novel Variants Detected by NGS in the ABCA4 Gene and Results of Analysis Using Bioinformatics Software

<table>
<thead>
<tr>
<th>Nucleotide Change</th>
<th>Protein</th>
<th>SpliceSite Finder-like</th>
<th>Gene Splicer</th>
<th>SpliceSite Finder-like</th>
<th>Gene Splicer</th>
<th>Average Difference</th>
<th>Polyphen</th>
<th>SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.91T&gt;C</td>
<td>p.W31R</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) W</td>
<td></td>
</tr>
<tr>
<td>c.184G&gt;T</td>
<td>p.P62S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) P</td>
<td></td>
</tr>
<tr>
<td>c.770T&gt;G</td>
<td>p.L257R</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Possibly damaging (0.308) m i F L</td>
<td></td>
</tr>
<tr>
<td>c.1255T&gt;C</td>
<td>p.F418S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) F</td>
<td></td>
</tr>
<tr>
<td>c.1531C&gt;T</td>
<td>p.R511C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (1.000) R</td>
<td></td>
</tr>
<tr>
<td>c.1745A&gt;G</td>
<td>p.N582S</td>
<td>0</td>
<td>0.74</td>
<td>0.82</td>
<td>77.8</td>
<td>Probably damaging (0.894) d K N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1868A&gt;G</td>
<td>p.Q623R</td>
<td>0</td>
<td>0.24</td>
<td>0</td>
<td>12.1</td>
<td>Probably damaging (0.957) Q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1964T&gt;G</td>
<td>p.F655C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) F</td>
<td></td>
</tr>
<tr>
<td>c.1977G&gt;A</td>
<td>p.M659I</td>
<td>0</td>
<td>0.75</td>
<td>0.85</td>
<td>79.8</td>
<td>Probably damaging (0.999) M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.2245G&gt;A</td>
<td>p.C748Y</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.928) g S A C</td>
<td></td>
</tr>
<tr>
<td>c.2410G&gt;A</td>
<td>p.A801T</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.98) A</td>
<td></td>
</tr>
<tr>
<td>c.2893A&gt;T</td>
<td>p.N965Y</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) N</td>
<td></td>
</tr>
<tr>
<td>c.3148G&gt;A</td>
<td>p.G1050S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Possibly damaging (0.786) G</td>
<td></td>
</tr>
<tr>
<td>c.3205A&gt;G</td>
<td>p.K1069E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.993) K</td>
<td></td>
</tr>
<tr>
<td>c.3270C&gt;A</td>
<td>p.D1093E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) D</td>
<td></td>
</tr>
<tr>
<td>c.3350C&gt;T</td>
<td>p.T1117I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.995) T</td>
<td></td>
</tr>
<tr>
<td>c.3655G&gt;C</td>
<td>p.A1219P</td>
<td>0.77</td>
<td>0.74</td>
<td>1.5</td>
<td>0.36</td>
<td>1.8</td>
<td>Probably damaging (0.991) A</td>
<td></td>
</tr>
<tr>
<td>c.3812A&gt;G</td>
<td>p.E1271G</td>
<td>0.8</td>
<td>0.35</td>
<td>0.71</td>
<td>21.8</td>
<td>0.14</td>
<td>Probably damaging (0.995) E</td>
<td></td>
</tr>
<tr>
<td>c.4177A&gt;G</td>
<td>p.V1393I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Benign (0.000) VI</td>
<td></td>
</tr>
<tr>
<td>c.4217A&gt;G</td>
<td>p.H1102G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.986) r p q a t k e g n S D H</td>
<td></td>
</tr>
<tr>
<td>c.4248C&gt;A</td>
<td>p.F1416L</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.27</td>
<td>0.07</td>
<td>Possibly damaging (0.374) a g d s T N</td>
<td></td>
</tr>
<tr>
<td>c.4326G&gt;A</td>
<td>p.N1442K</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Benign (0.047) p h l s n a c T Q K R</td>
<td></td>
</tr>
<tr>
<td>c.4467G&gt;T</td>
<td>p.R1489S</td>
<td>0.85</td>
<td>0.43</td>
<td>0.78</td>
<td>24.0</td>
<td>12.8</td>
<td>Possibly damaging (0.999) f W Y</td>
<td></td>
</tr>
<tr>
<td>c.4670A&gt;G</td>
<td>p.Y1557C</td>
<td>0.85</td>
<td>0.13</td>
<td>0.80</td>
<td>0.0</td>
<td>8.8</td>
<td>Probably damaging (0.999) f W Y</td>
<td></td>
</tr>
<tr>
<td>c.5138A&gt;G</td>
<td>p.Q1713R</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.997) Q</td>
<td></td>
</tr>
<tr>
<td>c.5177C&gt;A</td>
<td>p.T1726N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.880) s A T</td>
<td></td>
</tr>
<tr>
<td>c.5646G&gt;A</td>
<td>p.M1882I</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) M</td>
<td></td>
</tr>
<tr>
<td>c.6306C&gt;A</td>
<td>p.D2102E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.999) D</td>
<td></td>
</tr>
<tr>
<td>c.6718G&gt;2T&gt;C</td>
<td>p.T2240A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Probably damaging (0.991) T</td>
<td></td>
</tr>
<tr>
<td>c.160+2T&gt;C</td>
<td></td>
<td>0.81</td>
<td>0.86</td>
<td>0.79</td>
<td>0.0</td>
<td>44.4</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.160+2T&gt;C</td>
<td></td>
<td>0.82</td>
<td>0.81</td>
<td>0</td>
<td>0</td>
<td>81.5</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.1240+2A&gt;G</td>
<td></td>
<td>0.79</td>
<td>0.64</td>
<td>0</td>
<td>0</td>
<td>71.7</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.2382+1G&gt;A</td>
<td></td>
<td>0.9</td>
<td>0.92</td>
<td>0</td>
<td>0</td>
<td>90.9</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.2919+2A&gt;G</td>
<td></td>
<td>0.87</td>
<td>0.57</td>
<td>0.18</td>
<td>0.0</td>
<td>63</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.3522+9delG</td>
<td></td>
<td>0.9</td>
<td>0.89</td>
<td>0</td>
<td>0</td>
<td>89</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.3525+1G&gt;A</td>
<td></td>
<td>0.91</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>90.6</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.3814+2G&gt;A</td>
<td></td>
<td>0.74</td>
<td>0.82</td>
<td>0</td>
<td>0</td>
<td>78</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.4352+1G&gt;A</td>
<td></td>
<td>0.86</td>
<td>0.89</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
<td>New splice site 7 bp downstream</td>
<td></td>
</tr>
<tr>
<td>c.4655+1G&gt;T</td>
<td></td>
<td>0.86</td>
<td>0.89</td>
<td>0</td>
<td>0</td>
<td>87.5</td>
<td>New splice site 7 bp downstream</td>
<td></td>
</tr>
<tr>
<td>c.5312+1G&gt;A</td>
<td></td>
<td>0.81</td>
<td>0.91</td>
<td>0</td>
<td>0</td>
<td>86.1</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.5836+2A&gt;C</td>
<td></td>
<td>0.89</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
<td>88</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.6387+1G&gt;T</td>
<td></td>
<td>0.77</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
<td>82</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.6479+1G&gt;A</td>
<td></td>
<td>0.82</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
<td>85</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.6479+1G&gt;C</td>
<td></td>
<td>0.82</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
<td>56.6</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>c.1100+6T&gt;A</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td>0.93</td>
<td>91.6</td>
<td>Creates new splice site</td>
<td></td>
</tr>
</tbody>
</table>

For SpliceSiteFinder and GeneSplicer, 1 is the highest score for splice site activity and 0 is the lowest. For PolyPhen, 1 is the most damaging and 0 is the least.

For SIFT, tolerated amino acid residues are listed. Uppercase letters indicate tolerated amino acids; lowercase letters indicate less tolerated amino acids.

* Nucleotide positions and protein translation correspond to CCD5747.1 and NP_0003441.2, respectively.
STGD1 patients with the ABCA4 array started in 2000 (then with ~300 mutations); the number of mutations on the array has since doubled. The fact that only four mutations in all the array updates reoccurred shows that almost all new findings are rare. Three ABCA4 mutations in three different samples identified by sequencing—p.P1486L, p.G1961E, and p.R2106C—represented ABCA4 array false negatives. These three samples were screened many years ago when the quality of the ABCA4 array might have not been at its current level. One variant, c.768G>T/p.V256V, identified by the array was not confirmed by NGS; therefore, it represented the only NGS false negative since it was confirmed by Sanger sequencing. From 53 samples in which the combined APEX/NGS analysis had detected two mutations, only two samples carried the same two mutations—p.G1961E and p.R2149*. This proves once more the extraordinary heterogeneity of the ABCA4 alleles and the necessity of a cost-efficient full gene sequencing platform.

Fifteen benign ABCA4 missense variants were also detected (Supplementary Table S3, http://www.iovs.org/lookup/suppl/...).
Some with the c.5461–10T mutation in the sequenes remain unknown. A disease-associated mutation, although its functional conse-
c.4253...dbSNP (Supplementary Table S2, http://www.iovs.org/lookup/cohort) with no frequency data available even for those in 6T analysis (Table 2). As a result, one more intronic variant, c.1100 – no mutation after sequencing, were subjected to in silico anal-
TABLE 3.

<table>
<thead>
<tr>
<th>rs Number (if known)</th>
<th>Nucleotide Change</th>
<th>Protein</th>
<th>Heterozygotes (in 159 samples)</th>
<th>Homozygotes (in 159 samples)</th>
<th>MAF in dbSNP (if available)</th>
<th>MAF This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1801574</td>
<td>c.5682G&gt;C</td>
<td>p.L1894L</td>
<td>88</td>
<td>15</td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td>rs2275029</td>
<td>c.5844A&gt;G</td>
<td>p.P1948P</td>
<td>69</td>
<td>2</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>rs147957</td>
<td>c.5814A&gt;G</td>
<td>p.L1938L</td>
<td>69</td>
<td>6</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>rs1801555</td>
<td>c.6253T&gt;C</td>
<td>p.D2095D</td>
<td>51</td>
<td>6</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>rs1762114</td>
<td>c.6069T&gt;C</td>
<td>p.L2023I</td>
<td>27</td>
<td>3</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>rs1801359</td>
<td>c.6249C&gt;T</td>
<td>p.L2083I</td>
<td>21</td>
<td>1</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>rs1801666</td>
<td>c.4203C&gt;T</td>
<td>p.R1489S</td>
<td>18</td>
<td>2</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs417831</td>
<td>c.1206C&gt;T</td>
<td>p.H423H</td>
<td>17</td>
<td>—</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>rs77293072</td>
<td>c.6732G&gt;A</td>
<td>p.V256V</td>
<td>3</td>
<td>—</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>rs1801574</td>
<td>c.6253T&gt;C</td>
<td>p.L1894L</td>
<td>88</td>
<td>15</td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td>rs1801574</td>
<td>c.6253T&gt;C</td>
<td>p.L1894L</td>
<td>88</td>
<td>15</td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td>rs1801666</td>
<td>c.4203C&gt;T</td>
<td>p.R1489S</td>
<td>18</td>
<td>2</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs417831</td>
<td>c.1206C&gt;T</td>
<td>p.H423H</td>
<td>17</td>
<td>—</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>rs77293072</td>
<td>c.6732G&gt;A</td>
<td>p.V256V</td>
<td>3</td>
<td>—</td>
<td>0.09</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively.
variant was detected in three samples with only one disease-associated ABCA4 allele.

One of the proposed mechanisms for functional effect of synonymous variants is the introduction of less frequent codons resulting in delayed translation and folding of the protein. Examples of these mechanisms associated with a specific phenotype are described for other ABC transporters such as MDR1. The primary influence on codon use in mammals is the local base composition of the gene. Different codons can affect translation efficiency; it has been shown that G- and C-ending codons are more abundant in constitutive than in alternatively spliced exons in both Drosophila and humans.

From the 13 rare, silent ABCA4 variants detected in this study, six resulted in changes from a more frequent codon to a less frequent codon, and two of them, c.6342G>A/p.V2114V and c.6732G>A/p.V2244V (discussed earlier), represent a change from the most frequent codon to the least frequent codon (Table 4). Therefore, the c.6732G>A/p.V2244V change may affect the protein in a dual fashion by affecting splicing and by slowing the protein translation. Both variants were also found in samples with only one ABCA4 mutation. Because of the absence of family members for the four samples with the p.V2244V and p.V2114V variants, segregation analysis was not possible. Therefore, these two variants are classified as “suggestive” for disease association.

**DISCUSSION**

Next-generation sequencing of a large cohort of STGD1 patients, who had been screened previously with the ABCA4 array, discovered many new mutations in the coding region of ABCA4. NGS revealed 57 new disease-associated variants in 59 patients, and three more possibly pathogenic intronic and synonymous variants in another 11 patients. Therefore, though the (updated) array still remains a cost- and time-efficient first-pass screening tool, sequencing on an NGS platform would be a much more comprehensive approach. Determining the pathogenicity of rare missense (and also splice-affecting and synonymous) variants remains a significant challenge, especially for the ABCA4 gene, for which close to 700 possibly disease-associated variants have now been identified. Because more than half of these have been detected only once, the unequivocal classification for disease association, which is usually accomplished by segregation and functional analyses, presents an almost impossible task. Given that ABCA4-associated diseases are recessive, any given patient often represents the only affected member in a family with no siblings. Although determining the phase is often possible if parental samples are available, unequivocal segregation with the disease is often complicated or impossible. Functional analysis of the ABCA4 variants is also complicated because ABCA4 is expressed only in photoreceptors, which means no affected tissue can be obtained from patients, and because no direct functional test is available. Many frequent ABCA4 variants have been analyzed indirectly, such as by testing their effects on protein yield, folding, and ATP-binding and ATPase activity assays in mostly in vitro systems and Xenopus laevis models. Performing these experiments for hundreds of rare variants is unrealistic at this time.

In the present study the new variants were analyzed in a multiplex process. First, the frequent variants were removed from the analysis by filtering against public databases. Delterious mutations (nonsense and ins/del) were considered pathogenic based on their truncating or frameshift effect, or both, on the ABCA4 protein. Missense mutations were analyzed with well-known in silico predictive programs, PolyPhen and SIFT, which have shown to be approximately 80% reliable in correctly predicting functional variants. If family members were available, variants were analyzed for segregation with the disease in pedigrees (Fig. 2). Some variants were also screened in a large control cohort (364 persons) from our AMD studies, which includes ethnically matched controls older than 60 years of age without any retinal pathology as documented by thorough eye examination. However, given that all but seven new variants were found only once in 159 STGD1 patients, the screening of 364 controls had limited value. Rare synonymous changes in the ABCA4 ORF were analyzed for their effect on splicing with in silico programs and for codon use. All rare intronic variants were analyzed with splicing prediction programs.

Screening of the ABCA4 gene with any method is still far from 100% efficient. Even after complete sequencing of ABCA4 coding region in patients with definitive clinical diagnosis of STGD1, approximately 25% to 30% of patients remain with one identified pathogenic mutation, and no mutations were found in approximately 15% to 20%. The three most likely reasons for not finding all mutations were that a small subset (~1%) of patients harbored CNVs unde-

<table>
<thead>
<tr>
<th>Protein</th>
<th>Patient Count</th>
<th>Pathogenic ABCA4 Alleles in Those Patients</th>
<th>Codon Change</th>
<th>Codon Use Change in ABCA4</th>
<th>rs Number (if known)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.V2244V</td>
<td>3</td>
<td>All with 1 allele</td>
<td>GTG&gt; GTA</td>
<td>70&gt;14</td>
<td>rs77293072</td>
<td>From most frequent to least frequent</td>
</tr>
<tr>
<td>p.L988L</td>
<td>3</td>
<td>One with 2 alleles, Two with no alleles</td>
<td>CTC&gt; CTT</td>
<td>65&gt;33</td>
<td>rs61754034</td>
<td>From second frequent to third frequent</td>
</tr>
<tr>
<td>p.N343N</td>
<td>2</td>
<td>2 alleles, 1 allele</td>
<td>AAC&gt; AAG</td>
<td>36&gt;54</td>
<td>—</td>
<td>To most frequent</td>
</tr>
<tr>
<td>p.P47P</td>
<td>1</td>
<td>1 allele</td>
<td>CGG&gt; CCA</td>
<td>13&gt;44</td>
<td>rs4847281</td>
<td>From least frequent To most frequent</td>
</tr>
<tr>
<td>p.I171I</td>
<td>1</td>
<td>1 allele</td>
<td>ATC&gt; ATT</td>
<td>79&gt;43</td>
<td>—</td>
<td>From most frequent to less frequent</td>
</tr>
<tr>
<td>p.S206S</td>
<td>1</td>
<td>1 allele</td>
<td>AGC&gt; AGT</td>
<td>49&gt;16</td>
<td>—</td>
<td>From most frequent to less frequent</td>
</tr>
<tr>
<td>p.R500R</td>
<td>1</td>
<td>0 alleles</td>
<td>AGG&gt; AGA</td>
<td>29&gt;27</td>
<td>—</td>
<td>Neutral</td>
</tr>
<tr>
<td>p.A626A</td>
<td>1</td>
<td>1 allele</td>
<td>GGC&gt; GCA</td>
<td>10&gt;32</td>
<td>rs61754023</td>
<td>To more frequent</td>
</tr>
<tr>
<td>p.T1537T</td>
<td>1</td>
<td>0 alleles</td>
<td>AGC&gt; ACA</td>
<td>17&gt;32</td>
<td>—</td>
<td>To more frequent</td>
</tr>
<tr>
<td>p.A2022A</td>
<td>1</td>
<td>1 allele</td>
<td>AGC&gt; GGC</td>
<td>32&gt;10</td>
<td>—</td>
<td>To least frequent</td>
</tr>
<tr>
<td>p.S2072S</td>
<td>1</td>
<td>2 alleles</td>
<td>AGT&gt; AGG</td>
<td>16&gt;49</td>
<td>—</td>
<td>To most frequent</td>
</tr>
<tr>
<td>p.N2111N</td>
<td>1</td>
<td>2 alleles</td>
<td>AAC&gt; AAT</td>
<td>54&gt;36</td>
<td>—</td>
<td>To least frequent</td>
</tr>
<tr>
<td>p.V2114V</td>
<td>1</td>
<td>1 allele</td>
<td>GTG&gt; GTA</td>
<td>70&gt;14</td>
<td>rs61748520</td>
<td>From most frequent to least frequent</td>
</tr>
</tbody>
</table>
ected by PCR-based methods, a significant fraction of patho-
genetic mutations were outside the ABCA4 coding sequences, and some patients had diagnoses of ABCA4-associated dis-
cases that were phenocopies (diseases caused by mutations in
other known or yet to be discovered genes).

CNVs are predicted to be rare in the ABCA4 locus be-
cause several studies have found only a few cases (~1% of all
patients) with large (entire exon or chromosomesomal seg-
dent) deletions that avoid PCR-based detection methods. How-
ever, for complete mutational scanning, CNV analysis with
an array comparative genomic hybridization approach, or
with multiplex ligation-dependent probe amplification,
could be included.

The present study also determined that many pathogenic
mutations are likely located outside the ABCA4 ORF because
the second mutation, required for the genetic diagnosis of
STGD1, was not found in approximately half of all patients
with one mutation. Although one could argue that some of
these patients could be carrying the ABCA4 variant by
chance because of the high population frequency of ABCA4
variants (estimated 1:20), it is highly unlikely that a patient
with a clinical diagnosis of STGD1 and carrying one muta-
ton does not have the second pathogenic variant. Detection
of disease-associated variants outside the ABCA4 coding
sequences will be accomplished by sequencing of the entire
130-kb ABCA4 genomic locus in patients with one identified
mutation, a study that is in progress.

Finally, in patients in whom no disease-associated vari-
ants are found, whole exome or genome approaches can be
used to determine new gene mutations that cause STGD-
mimicking phenotypes. This approach can be preceded by
sequencing known genes, such as RDS/PRPH2 (gene for
multifocal pattern dystrophy, 3 exons),27,28 ELOVL4 (domi-
nant STGD-like disease gene, 6 exons),29,30 VMD2 (Best
disease gene [recessive forms resemble STGD], 11 exons),
RS1 (retinoschisis gene, 6 exons),31 and CNGB3 (achroma-
topsia gene, 17 exons).32 In our studies, however, we have
not found disease-associated mutations in RDS or ELOVL4
genes in 30 to 40 STGD patients with no mutations in
ABCA4; therefore, the yield by this approach is expected to
be limited.

The fraction of genocopies (i.e., clinical misdiagnoses) at
a given clinic depends primarily on the depth of clinical
analyses. At our centers patients have primarily undergone
detailed clinical work-up with all the techniques shown in
Materials and Methods; therefore, the fraction of genocopies
is expected to be small. However, in an average clinical,
diagnosis is based mainly on ophthalmoscopic examination
(fundus photography) with only a few additional techniques
(e.g., BCVA, OCT, microperimetry) yielding less stringent
criteria for final diagnosis. Moreover, even exceptionally
extensive clinical data are often not enough for pinpointing
the possible genetic cause. As a reminder, depending on the
severity of the ABCA4 mutation and the stage of the disease
diagnosed, ABCA4-associated pathology presents in a wide
range of phenotypes from mild fundus flavimaculatus to
CRD and even RP-like phenotypes. The latter two pheno-
types are caused by tens of distinct genes.

Given the substantially overlapping phenotypes and sev-
eral treatment options currently in late stages of preclinical
development or in clinical trials, the correct and compre-
hensive molecular diagnosis of ABCA4-associated diseases
is crucial. The NGS platform is a time- and cost-efficient tool
to analyze large and variable genes simultaneously in large
cohorts and could be used for diagnostic applications.

References

ATP-binding transporter gene (ABCR) is mutated in recessive Star-
retinitis pigmentosa and cone-rod dystrophy caused by splice site
the ABC4 (ABCR) gene are the major cause of autosomal recessive
caused by a homozygous mutation in the Stargardt disease gene
5. Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR. Null missense ABCR
(ABCA4) mutations in a family with Stargardt disease and retinitis pig-
6. Allikmets R. Stargardt disease: from gene discovery to therapy. In:
Tophoven-Tink J, Barnstable CJ, eds. Retinal Degenerations: Biol-
ogy, Diagnostics and Therapeutics. Totowa, NJ: Humana Press;
2007:105–118.
mutation in the ABCR gene is a mild frequent founder mutation in
the Western European population and allows the classification of
variation in the ABC4 (ABCR) gene in Stargardt disease and age-related
9. Lewis RA, Shroyer NF, Singh N, et al. Genotype/phenotype analysis of
a photoreceptor-specific ATP-binding cassette transporter gene,
10. Shroyer NF, Lewis RA, Yatsenko AN, Wensel TG, Lupski JR. Coseg-
gregation and functional analysis of mutant ABCR (ABCA4) alleles in
families that manifest both Stargardt disease and age-related macular
11. Allikmets R. Simple and complex ABCR: genetic predisposition to
13. Ernest PJ, Boon CJ, Klevering BJ, Hoefsloot LH, Huygen CB. Out-
come of ABCA4 microarray screening in routine clinical practice.
mutation analysis of the ABCA4 (ABCR) gene in autosomal rece-
15. Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. Late-onset Star-
gardt disease is associated with missense mutations that map
outside known functional regions of ABCR (ABCA4). Hum Genet.
16. Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. An ABCA4
genomic deletion in patients with Stargardt disease. Hum Mutat.
17. Kong J, Kim SR, Binley K, et al. Correction of the disease pheno-
type in the mouse model of Stargardt disease by lentiviral gene
molecule RPE65 antagonists limit the visual cycle and prevent
19. Chen R, Davydov EV, Sirota M, Butte AJ. Non-synonymous and
synonymous coding SNPs show similar likelihood and effect size of
Silent (synonymous) SNPs: should we care about them? Methods Mol.
22. Kilman RM, Bernal CA. Unusual usage of AGG and TTG codons in


