Detailed genetic characteristics of an international large cohort of patients with Stargardt disease: ProgStar study report 8

Kaoru Fujinami,1,2,3,4 Rupert W Strauss,3,4,5,6,7 John (Pei-Wen) Chiang,8 Isabelle S Audo,9,10 Paul S Bernstein,11 David G Birch,12 Samantha M Bomotti,5 Artur V Cideciyan,13 Ann-Margret Ervin,5 Meghan J Marino,14 José-Alain Sahel,10,15,16 Saddek Mohand-Said,9,10 Janet S Sunness,17 Elias I Traboulsi,14 Sheila West,5 Robert Wojciechowski,5 Eberhart Zrenner,18,19 Michel Michaelides,3,4 Hendrik P N Scholl,20,21 ProgStar Study Group, On behalf of the ProgStar Study Group

ABSTRACT
Background/aims To describe the genetic characteristics of the cohort enrolled in the international multicentre progression of Stargardt disease 1 (STGD1) studies (ProgStar) and to determine geographic differences based on the allele frequency.

Methods 345 participants with a clinical diagnosis of STGD1 and harbouring at least one disease-causing ABCA4 variant were enrolled from 9 centres in the USA and Europe. All variants were reviewed and in silico analysis was performed including allele frequency in public databases and pathogenicity predictions. Participants with multiple likely pathogenic variants were classified into four national subgroups (USA, UK, France, Germany), with subsequent comparison analysis of the allele frequency for each prevalent allele.

Results 211 likely pathogenic variants were identified in the total cohort, including missense (63%), splice site alteration (18%), stop (9%) and others. 50 variants were novel. Exclusively missense variants were detected in 139 (50%) of 279 patients with multiple pathogenic variants. The three most prevalent variants of these patients with multiple pathogenic variants were classified into four national subgroups (USA, UK, France, Germany), with subsequent comparison analysis of the allele frequency for each prevalent allele.

Conclusions There is a large spectrum of ABCA4 sequence variants, including 50 novel variants, in a well-characterised cohort thereby further adding to the unique allelic heterogeneity in STGD1. Approximately half of the cohort harbours missense variants only, indicating a relatively mild phenotype of the ProgStar cohort. There are significant differences in allele frequencies between nations, although the three most prevalent variants are shared as frequent variants.

INTRODUCTION
Stargardt disease 1 (STGD1; MIM 248200) is the most prevalent inherited macular dystrophy, which is an autosomal recessive condition caused by pathogenic sequence variants in the ABCA4 gene (ATP-binding cassette subfamily A member 4; MIM 601691).1 2 ABCA4 encodes the retinal-specific transmembrane protein and is involved in the active transport of retinoids in visual cycle.1 2 Patients with STGD1 typically presents with bilateral central visual loss, including central scotoma and reduced visual acuity, and with characteristic macular atrophy surrounded by yellow-white flecks at the level of the retinal pigment epithelium.1 3–6 Highly variable phenotypes, severity and progression of STGD1 have been found in ABCA4-associated retinopathy.5–16

There is also a very high allelic heterogeneity in ABCA4, with more than 1000 sequence variations reported to date.1 17–21 The phenotypic variability and the genetic heterogeneity pose marked challenges in attempts to establish genotype–phenotype correlations of ABCA4-associated retinopathy. However, comprehensive clinical and genetic investigations of STGD1 in a large cohort based on well-established eligibility criteria are lacking and would likely help to address the aforementioned challenges. Therefore, the international multicentre ‘Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar)’ studies were established.3

The purpose of the present study is to describe the detailed genetic characteristics of the large STGD1 cohort enrolled into the ProgStar studies. This study also provides an opportunity to determine geographic differences in the allele frequency of prevalent ABCA4 variants.

MATERIAL AND METHODS
Patients
In ProgStar, patients with STGD1 were enrolled from nine centres in the USA and Europe: The Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland (JHU); Greater Baltimore Medical Centre, Baltimore, Maryland (GBMC); Scheie Eye Institute, University of Philadelphia, Philadelphia, Pennsylvania (Penn); Retina Foundation of the Southwest, Dallas, Texas (RFSW); Moran Eye Centre, Salt Lake City, Utah (MEC); Cole Eye Institute, Cleveland Clinic, Cleveland, Ohio (CC); Moorfields Eye Hospital, London, UK (MEH, UK); Université de Paris 06, Institut national de la santé et de la recherche médicale, Paris, France (INSERM, France); and Eberhard-Karls University
The main clinical/molecular genetic eligibility criteria were as follows: (1) patients (at least 6 years old) with at least two ABCA4 variants or one ABCA4 variant associated with a typical STGD1 phenotype and (2) presence of a well-defined atrophic lesion with/without flecks at the most recent visit of at least 300 μm in diameter (the total area of all lesions <12 mm²). A fundus autofluorescence image of a representative patient enrolled in the ProgStar studies is shown in online supplementary figure 1.

Participants with available clinical and genetic information from the two ProgStar study cohorts (ProgStar retrospective and prospective studies) were included, with one affected proband from each family selected. Three hundred and forty-five participants with a clinical diagnosis of STGD1 and harbouring at least one ABCA4 variant were ascertained. The protocols of the ProgStar retrospective and prospective studies adhered to the provisions of the Declaration of Helsinki and were approved by the local ethics committee of all participating institutions.

**Mutation detection and in silico molecular genetic analysis**

ABCA4 gene screening was performed in all participants (n=345) between 2000 and 2014 with the following strategies: PCR enrichment based targeted next-generation sequencing (NGS) (n=143), gene chip array (n=44), single-strand conformation polymorphism (n=24) and direct sequencing (n=134). All the detected variants were confirmed by direct sequencing, and cosegregation analysis was performed in 28 families.

Pathogenicity of all the detected variants was analysed using two public databases reporting allele frequencies in the general population (The ExAC Browser and 1000 Genomes Browser) and four software prediction programs (SIFT, PolyPhen2, Mutation Taster and HSF3.0).

All variants located within all 50 exons and exon–intron boundaries (<±11 base pairs) were classified as ‘likely’ pathogenic if they met the following criteria modified according to the previous reports: (1) allele frequency of less than 0.3% (calculated out based on the expected carrier frequency of 1/20–1/50) for all variants except for the 10 most prevalent variants and 1 variant (c.5603A>T, p.Asn1868Ile) which have proven pathogenicity and high frequency; (2) pathogenicity in missense variants confirmed by at least two of the three prediction programs (PolyPhen2, SIFT, mutation taster); (3) significant splicing effect in intronic and synonymous exonic variants confirmed by HSF3.0. All variants within the exons and exon–intron boundaries that did not meet the criteria were classified as ‘less likely’ pathogenic. ‘Deep’ intronic variants (>10 base pairs distant from the end of exon) were predicted as uncertain. Variants that are usually found in a complex with a common likely pathogenic variant were also predicted as ‘uncertain’. ‘Uncertain’ variants were classified as ‘less likely’ pathogenic for the purpose of this analysis.

**Genotype group classification**

Patients harbouring at least two likely pathogenic variants were classified into three genotype groups based on the severity of predicted mutational damage: group A: patients with two or more severe/null variants; group B: patients with one severe/null variant and at least one missense or in-frame insertion/deletion and group C: individuals with two or more missense or in-frame insertion/deletion variants. Severe/null variants were those that would be expected to affect splicing or to introduce a premature truncating codon in the protein if translated: stop, frame shift, intron variants in splice regions with significant splice site alteration; exonic synonymous variants with significant splice site alteration; and nonsense variants with significant splice site alteration (eg, nucleotide change at the start/end of exon).

**Subgroup analysis for nation, institution and sex**

In order to investigate differences of geographical location, institutions and sex, subgroup analyses were performed in patients with at least two likely pathogenic variants by comparing the allele frequency of prevalent pathogenic variants. The prevalent likely pathogenic variants were defined as variants with an allele frequency of at least 2.0% in each subgroup or likely pathogenic variants of at least 1.5% in the total cohort of patients harbouring at least two likely pathogenic variants. An isolated variant identified only once in each subgroup was excluded even when the frequency met the criteria above.

Geographical subgroup analysis was performed among the four nations (USA, UK, France, Germany) and regional analysis among the six institutions in the USA (JHU, GBMC, PENN, RFSW, MEC, CC), respectively. Gender analysis was performed to compare the allele frequency of prevalent likely pathogenic variants between females and males.

Statistical association between each subgroup and presence of each prevalent allele was investigated with categorical testing of the independence (Fisher’s exact test/χ² test) using commercially available software: Excel Tool 2015 (Social Survey Research Information, Tokyo, Japan). P values less than 0.05 were considered to indicate statistical significance.

**RESULTS**

**Clinical findings**

The clinical findings of the total cohort of 345 unrelated probands with STGD1 are summarised in table 1. The cohort included 150 patients from the USA (JHU-23, GBMC-27, PENN-25, RFSW-33, MEC-21, CC-21), 85 from UK (MEH), 61 from France (INSERM) and 49 from Germany (EKU): 191 females and 154 males. The median age of onset (defined as the age at which any symptom was first noted by the patient) was 19.0 years (range 4–68 years), and the median age at baseline examination was 28.0 years (range 7–71 years). The median equivalent logarithm of the minimum angle of resolution visual acuities of the right eye and left eye were 0.80 (range 0.10–1.70; equivalent to 20/25 to 20/1000) and 0.80 (range 0.10–1.56 equivalent to 20/25 to 20/720), respectively.

**Detected variants and results of in silico molecular genetic analysis**

The genetic findings of the total cohort are summarised in online supplementary table 1. Two hundred and forty-five patients harboured at least two likely pathogenic variants.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of clinical findings in 345 patients with STGD1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age of onset (years)</td>
<td>19.0 (4–68)</td>
</tr>
<tr>
<td>Median age at examination (years)</td>
<td>28.0 (7–71)</td>
</tr>
<tr>
<td>Median duration of disease (years)</td>
<td>5.0 (0–36)</td>
</tr>
<tr>
<td>Median LogMAR visual acuity (right eye)</td>
<td>0.8 (–0.10 to 1.70)</td>
</tr>
<tr>
<td>Median LogMAR visual acuity (left eye)</td>
<td>0.8 (–0.10 to 1.56)</td>
</tr>
</tbody>
</table>

The median age of onset was defined as the age at which any symptom was first noted by the patient. The duration of disease was calculated as the difference between age of onset and age at the baseline examination.

The detailed findings of fundus autofluorescence are described elsewhere. LogMAR, logarithm of the minimum angle of resolution; STGD1, Stargardt disease 1.
variants were identified in the total cohort: including missense variants (n=153), splice site alteration (n=45), stop (n=19), frame shift (n=18), deep intronic variants (n=7), large exonic deletion (n=1), in-frame deletion (n=1) and duplication (n=1) (figure 1A). In silico molecular analysis detected 211 likely pathogenic variants and 34 less likely pathogenic variants (online supplementary table 1). The detailed results of in silico molecular genetic analysis are presented in online supplementary table 2.

Overall, there were 279 patients with multiple (at least two) pathogenic variants, 62 with one pathogenic variant and 4 with only less likely pathogenic variants. The 211 likely pathogenic variants identified in the total cohort included missense variants (n=133), splice site alteration (n=38), stop (n=19), frame shift (n=18), large exonic deletion (n=1), in-frame deletion (n=1) and duplication (n=1) (figure 1B). There were three missense variants with significant splice site alteration and one synonymous variant with significant splice site alteration (c.1A>G, p.Met1Val; c.1760G>A, p.Arg587Lys; c.4538A>G, p.Gln1513Arg; c.768G>T, p.Val256Val). Thirty-four less likely pathogenic variants included twenty missense variants, seven uncertain deep intronic variants, four synonymous variants with no significant effect on protein damage or splice site alteration and three variants in splice regions with no significant predicted effect on splicing (online supplementary table 2). Thirteen of the less likely pathogenic variants had high allele frequency (>0.1% in general population), most of which were detected before the establishment of a public database of allele frequency in the general population.

Fifty novel likely pathogenic variants were identified in the total cohort, including missense variants (n=27), splice site alteration (n=12), frame shift (n=7), stop (n=3) and duplication (n=1) (online supplementary tables 1 and 2, figure 1C). These novel variants were detected by PCR enrichment based targeted NGS (n=26), direct sequencing (n=25) and gene chip array (n=1). Two variants were identified both by PCR enrichment based targeted NGS and direct sequencing and one variant was identified by direct sequencing and gene chip array, respectively.

Three pairs of co-inherited complex variants were identified in 27 patients in the total cohort: c.1622T>C, p.Leu541Pro and c.3113C>T, p.Ala1038Val (n=18); c.4222T>C, p.Trp1408Arg and c.4918C>T, p.Arg1640Trp (n=5); and c.3758C>T, p.Thr1253Met and c.5882G>A, p.Gly1961Glu (n=4) (online supplementary tables 1 and 2). Two variants were predicted as less likely pathogenic (c.3113C>T, p.Ala1038Val and c.3758C>T, p.Thr1253Met). It is uncertain which of the two, or arguably less likely whether both variants, p.Trp1408Arg and/or p.Arg1640Trp, are pathogenic, so both were classified as likely pathogenic in this analysis (online supplementary tables 1 and 2). 12

Genotype groups

Two hundred and seventy-nine patients harbouring two or more likely pathogenic variants were classified into three genotype groups: group A=only severe/null variant (n=16); group B=one severe/null variant and at least one missense or in-frame insertion/deletion (n=124) and group C=two or more missense or in-frame insertion/deletion variant (n=139) (figure 1D).

Subgroup analysis for nation, institution and sex

There were 279 patients harbouring two or more likely pathogenic variants in total, with 121 patients from the USA (JHU, GBMC-25, PENN-24, RFSW-24, MEC-12, CC-15), 71 from UK (MEH), 52 from France (INSERM) and 35 from Germany (EKU). There were 10 prevalent variants in this cohort (table 2, online supplementary figure 2). The three most prevalent variants in total cohort were p.G1961E (15.1%), p.G863A (7.2%) and c.5461–10 T>C (4.8%), with high allele frequency of at least 2.8% in the four nations.

Geographical subgroup comparison analysis was performed between the four nations in terms of the 23 most prevalent likely pathogenic variants identified both in the total cohort and in each geographical subgroup. There was a statistically significant difference in nine variants between nations: (1) c.1317G>A, p.Trp439Ter and c.1906C>T, p.Gln638Ter (prevalent in UK); (2) c.3386G>T, p.Arg1129Leu; c.3603A>T, p.Asn1868Ile; c.2041C>T, p.Arg681Ter (prevalent in France); (3) c.1648G>A, p.Gly550Arg; c.183G>C, p.Met61Ile; c.6112C>T, p.Arg2038Trp; c.6721C>G, p.Leu2241Val (prevalent in Germany). No variants prevalent in the USA had a statistically significant difference in prevalence compared with the other three nations (table 3, figure 2).

Regional subgroup analysis was performed between the six institutions (JHU, GBMC, PENN, RFSW, MEC, CC) in the USA. A statistically significant difference was found for five variants: c.6079C>T, p.Leu2027Phe (prevalent in JHU); c.5395A>G, p.Asn1799Asp (prevalent in RFSW); c.4253+4C>T, p.Glu2383X; c.3259G>A, p.Glu1087Lys and c.1605G>A, p.Arg535Val (prevalent in MEC).

DISCUSSION

The broad spectrum of ABCA4 variants was documented in a well-characterised large cohort with STGD1 based on well-established inclusion and exclusion criteria. Three hundred and forty-five unrelated probands with STGD1 harbour 245 specific ABCA4 variants, including 211 likely pathogenic and 34 less likely pathogenic variants. To our knowledge, the present cohort is the largest among STGD1 studies and provides data on the distribution and prevalence of these ABCA4 variants.

A broad range of variants was distributed throughout the ABCA4 gene. 62% of the variants were missense mutations in coding regions, with intronic variants (15%) located in exon–intron boundaries (<±11 base pairs) and deep intronic variants (>10 base pairs) (3%) also detected (online supplementary tables 1 and 2, figure 1A). These findings confirm that there are no specific mutation hot spots in ABCA4; hence comprehensive genetic screening is recommended for mutation detection.1 The possibility of missing large exonic deletions or insertion/deletions of over 10 nucleotides raises the potential validity of applying whole-genome sequencing in the future which would enable the identification of such alterations.

In silico molecular genetic analysis revealed 211 likely pathogenic and 34 less likely pathogenic variants in the present study. These 34 variants are composed of 19 missense variants with no significant protein damage, 1 missense variant with uncertain effect, 7 deep intronic variants with uncertain effect, and 5 synonymous exonic variants and 2 variants in splice region, both of which had no significant effect on splicing. Due to the
Figure 1 The distribution of variants and genotype groups of cases with Stargardt disease 1 (STGD1) recruited to the Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) studies. (A) Distribution of 245 variants detected in the total cohort of 345 patients: 245 variants were identified in the total cohort including missense variants (n=152), intronic variants in splice regions with predicted splice site alteration (n=34), stop (n=19), frame shift (n=18), deep intronic variants with uncertain effect (n=7), exonic synonymous with potential splice site alteration (n=4), intronic variants in splice region with potential splice site alteration (n=3), missense variants with significant splice site alteration (n=3), exonic synonymous variants with significant splice site alteration (n=1), missense variants with uncertain effect (n=1), large exonic deletion (n=1), in-frame deletion (n=1) and duplication (n=1). (B) Distribution of 211 likely pathogenic variants: 211 likely pathogenic variants comprising missense variants (n=133), intronic variants in splice regions with significant splice site alteration (n=34), stop (n=19), frame shift (n=18), missense variants with significant splice site alteration (n=3), exonic synonymous variants with significant splice site alteration (n=1), large exonic deletion (n=1), in-frame deletion (n=1) and duplication (n=1). (C) Distribution of 50 novel likely pathogenic variants: 50 likely pathogenic variants were novel, including missense variants (n=27), intronic variants in splice regions with significant splice site alteration (n=12), frame shift (n=7), stop (n=3) and duplication (n=1). (D) Distribution of genotype groups in 279 patients with multiple likely pathogenic variants: 279 patients harbouring multiple likely pathogenic variants were classified into three genotype groups based on the severity of predicted mutational damage: genotype group A with two or more severe/null variants (n=16), genotype group B with one severe/null variant and at least one missense or in-frame insertion/deletion variant (n=124) and genotype group C with two or more missense or in-frame insertion/deletion variants (n=139).
variants. A high allele frequency (>0.1%) was revealed in evaluated, although there have been several reports suggesting intronic and exonic synonymous variants were not extensively inherent limitations of prediction protocols, the effects of deep with high allele frequency (4.46%) has been classified as likely general population. One variant (c.5603A>T, p.Asn1868Ile) to the availability of data of allele frequency in the normal/

### Table 2  Ten prevalent variants in 279 patients with multiple likely pathogenic variants

<table>
<thead>
<tr>
<th>Nucleotide change, amino acid change/effect</th>
<th>Allele frequency in total ProgStar cohort with multiple likely pathogenic variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.5882G&gt;A, p.Gly1961Glu</td>
<td>15.05%</td>
</tr>
<tr>
<td>c.2558G&gt;C, p.Gly863Ala</td>
<td>7.17%</td>
</tr>
<tr>
<td>c.5461–101T&gt;C, splice site alteration</td>
<td>4.84%</td>
</tr>
<tr>
<td>c.4139C&gt;T, p.Pro1380Leu</td>
<td>3.94%</td>
</tr>
<tr>
<td>c.1622T&gt;C, p.Leu641Pro</td>
<td>2.69%</td>
</tr>
<tr>
<td>c.5714+5G&gt;A, splice site alteration</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.3322C&gt;T, p.Arg1108Cys</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.6079C&gt;T, p.Leu2027Phe</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.6320G&gt;A, p.Arg2107His</td>
<td>1.61%</td>
</tr>
<tr>
<td>c.6089G&gt;A, p.Arg2030Gln</td>
<td>1.61%</td>
</tr>
</tbody>
</table>

**Table 3  Geographical subgroup analysis of allele frequency in 23 prevalent likely pathogenic variants between four nations**

<table>
<thead>
<tr>
<th>Nucleotide change, amino acid change/effect</th>
<th>USA</th>
<th>UK</th>
<th>France</th>
<th>Germany</th>
<th>ProgStar</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.5882G&gt;A, p.Gly1961Glu</td>
<td>13.60%</td>
<td>16.20%</td>
<td>13.55%</td>
<td>20.00%</td>
<td>15.05%</td>
</tr>
<tr>
<td>c.2558G&gt;C, p.Gly863Ala</td>
<td>7.00%</td>
<td>9.90%</td>
<td>2.88%</td>
<td>8.60%</td>
<td>7.17%</td>
</tr>
<tr>
<td>c.5461–101T&gt;C, splice site alteration</td>
<td>5.00%</td>
<td>4.90%</td>
<td>2.88%</td>
<td>7.10%</td>
<td>4.84%</td>
</tr>
<tr>
<td>c.4139C&gt;T, p.Pro1380Leu</td>
<td>4.10%</td>
<td>4.90%</td>
<td>4.81%</td>
<td>0.00%</td>
<td>3.94%</td>
</tr>
<tr>
<td>c.1622T&gt;C, p.Leu541Pro</td>
<td>3.30%</td>
<td>0.70%</td>
<td>1.92%</td>
<td>5.70%</td>
<td>2.69%</td>
</tr>
<tr>
<td>c.5714+5G&gt;A, splice site alteration</td>
<td>1.70%</td>
<td>1.40%</td>
<td>2.88%</td>
<td>5.70%</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.3322C&gt;T, p.Arg1108Cys</td>
<td>1.70%</td>
<td>2.80%</td>
<td>2.88%</td>
<td>2.90%</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.6079C&gt;T, p.Leu2027Phe</td>
<td>1.70%</td>
<td>3.50%</td>
<td>3.85%</td>
<td>0.00%</td>
<td>2.33%</td>
</tr>
<tr>
<td>c.6320G&gt;A, p.Arg2107His</td>
<td>2.10%</td>
<td>0.70%</td>
<td>1.92%</td>
<td>1.40%</td>
<td>1.61%</td>
</tr>
<tr>
<td>c.6089G&gt;A, p.Arg2030Gln</td>
<td>1.20%</td>
<td>2.80%</td>
<td>0.96%</td>
<td>1.40%</td>
<td>1.61%</td>
</tr>
<tr>
<td>c.3386G&gt;T, p.Arg1129Leu*</td>
<td>0.80%</td>
<td>0.00%</td>
<td>4.81%</td>
<td>1.40%</td>
<td>1.43%</td>
</tr>
<tr>
<td>c.4577C&gt;T, p.Thr1526Met</td>
<td>2.10%</td>
<td>0.70%</td>
<td>1.92%</td>
<td>0.00%</td>
<td>1.43%</td>
</tr>
<tr>
<td>c.4469G&gt;A, p.Cys1490Tyr</td>
<td>0.80%</td>
<td>3.50%</td>
<td>0.96%</td>
<td>0.00%</td>
<td>1.43%</td>
</tr>
<tr>
<td>c.5603A&gt;T, p.Asn1868Ile*</td>
<td>0.00%</td>
<td>0.00%</td>
<td>4.81%</td>
<td>0.00%</td>
<td>0.90%</td>
</tr>
<tr>
<td>c.2041C&gt;T, p.Arg681Ter*</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3.85%</td>
<td>1.40%</td>
<td>0.90%</td>
</tr>
<tr>
<td>c.3364G&gt;A, p.Glu1122Lys</td>
<td>0.40%</td>
<td>2.10%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.72%</td>
</tr>
<tr>
<td>c.6088C&gt;T, p.Arg2030Ter</td>
<td>0.40%</td>
<td>2.10%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.72%</td>
</tr>
<tr>
<td>c.1648G&gt;A, p.Gly550Arg*</td>
<td>0.40%</td>
<td>0.00%</td>
<td>2.90%</td>
<td>0.54%</td>
<td></td>
</tr>
<tr>
<td>c.1317G&gt;A, p.Thr439Ter*</td>
<td>0.00%</td>
<td>2.10%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.54%</td>
</tr>
<tr>
<td>c.1906C&gt;T, p.Gln636Ter*</td>
<td>0.00%</td>
<td>2.10%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.54%</td>
</tr>
<tr>
<td>c.183G&gt;C, p.Met611Le*</td>
<td>0.00%</td>
<td>0.00%</td>
<td>2.90%</td>
<td>0.36%</td>
<td></td>
</tr>
<tr>
<td>c.6112C&gt;T, p.Arg2038Tfp*</td>
<td>0.00%</td>
<td>0.00%</td>
<td>2.90%</td>
<td>0.36%</td>
<td></td>
</tr>
<tr>
<td>c.6721C&gt;G, p.Leu2241Val*</td>
<td>0.00%</td>
<td>0.00%</td>
<td>2.90%</td>
<td>0.36%</td>
<td></td>
</tr>
</tbody>
</table>

The high allele frequency show on grey background was defined as the allele frequency of at least 2.0% in each subgroup and the allele frequency of at least 1.5% in the total ProgStar cohort.

Two hundred and seventy-nine patients harbouring multiple likely pathogenic consists of 121 patients from the USA, 71 from UK, 52 from France and 35 from Germany.

*Comparison analysis revealed statistical difference in nine variants.

The prevalent likely pathogenic variants were defined as variants with allele frequency of at least 1.5% in 279 patients with multiple likely pathogenic variants.

Pathogenic in this study following a recent report supporting disease causation. However, it is challenging to calculate the actual allele frequency of this variant in a disease cohort considering the often applied filtration method of excluding variants with a frequency of >1%. This illustrates the importance of continuously revisiting the pathogenicity ascribed to variants based on evolving genetic and clinical data especially where establishing disease causation is inherently more challenging; for example, for missense and deep intronic sequence variants.

Fifty novel pathogenic variants were identified with half of these detected by PCR enrichment based targeted NGS (online supplementary tables 1 and 2, figure 1C). This finding, and our observation that a significant number of these novel variants were intronic, supports that PCR enrichment based targeted NGS is currently one of the most useful and comprehensive methods to detect previously reported plus novel variants in ABCA4 and that additional novel variants remain to be identified, thereby making ABCA4 arguably one of the most allelic heterogeneous genes in human genetics. On the other hand, the conventional sequencing methods such as gene chip array, single-strand conformation polymorphism and direct sequencing were also useful to detect prevalent pathogenic variants mostly published (online supplementary table 2).

Genotype group classification was performed in 279 patients harbouring two or more likely pathogenic variants: 5.7% (16/279) in group A, 44.4% (124/279) in group B and 49.8% (139/279) in group C (figure 1D). In previous studies of ABCA4-associated retinopathy, the proportion of genotype groups A, B and C was 1.6% (3/64), 40.6% (26/64) and 54.7% (35/64) in an adult cohort (age of onset >17 years) and 20.6% (7/34), 44.1% (15/34) and 35.2% (12/34) in a paediatric cohort.

The high allele frequency show on grey background was defined as the allele frequency of at least 2.0% in each subgroup and the allele frequency of at least 1.5% in the total ProgStar cohort.

Two hundred and seventy-nine patients harbouring multiple likely pathogenic consists of 121 patients from the USA, 71 from UK, 52 from France and 35 from Germany.
Approximately half of the ProgStar cohort harbours only missense variants and this proportion was similar to that of the aforementioned previously reported adult cohort (milder disease), in contrast to the paediatric cohort (severe disease). These findings are in keeping with the relatively mild phenotype of the ProgStar cohort, since the inclusion criteria of well-demarcated macular atrophy does not include patients with very severe phenotypes such as extended atrophy associated with more severe ABCA4 variants.

In addition, the high prevalence (15.1%) of the most common variant (p.G1961E) is in keeping with the phenotypic subset (macular atrophy with/without flecks) selected for ProgStar studies. It is well-established that the clinical effect of homozygous p.G1961E includes adult-onset disease, atrophy confined to the macula, retinal dysfunction confined to macular and variable visual acuity.\(^7\)\(^8\)\(^11\) In previous studies, the prevalence of this variant (p.G1961E) was 12.5% in an adult cohort and 5.9% in a paediatric cohort, respectively.\(^5\) Therefore, the proportion of patients with this mild allele, p.G1961E should be one of the main factors that predicts disease severity of the cohort. Additional detailed genotype–phenotype association/correlations studies based on each genotype and each allele are needed to understand the disease mechanism of ABCA4-associated retinopathy.

Geographical subgroup analysis between the four nations revealed significant differences in nine variants, while three variants (c.5582G>A, p.Gly1961Glu; c.2588G>C, p.Gly863Al and c.5461–10T>C, splice site alteration) were frequently found in all four nations. There was also a significant regional difference between the six participating institutions in the USA with respect to five prevalent variants. These findings provide preliminary data suggesting relatively unique genetic backgrounds of geographic areas/institutions especially regarding prevalent variants. Additional studies using haplotype analyses of whole-genome sequence results would be helpful to elucidate founder effects associated with ethnicity, which should underlie some of the geographical/regional differences.

There are several limitations in this study with regards to the gene screening and gene analysis methods, as well as the small number of families where cosegregation was possible, which may partly relate to the study design but are in keeping with the vast majority of inherited retinal disease studies. In the ProgStar cohort, deep intronic, synonymous and copy number variants were not screened for and analysed due to the limitation of the applied screening/analysis methods, including the conventional target direct screening of 50 exons and exon–intron boundaries. In addition, the possible presence of causative/modifier variants outside of the ABCA4 genes remains to be evaluated.
Moreover, clinical effects of common or rare ‘benign’ variants in cis or in trans need to be considered in the ABCA4 gene, especially in light of the recently identified variant (c.5603A>T, p.Asn1868Ile) which has a high allele frequency. Therefore, more advanced and comprehensive screening/analysis techniques using newly developed sequencing method, prediction tools and public databases including whole-genome sequencing, combined annotation-dependent depletion and genome aggregation database (gnomAD) would help to obtain the ‘conclusive’ molecular genetic diagnosis in a greater proportion of patients in this ProgStar cohort. Furthermore, the ProgStar studies have focused on a phenotypic subset (macular atrophy with/without flecks which can be tracked over time) and expanded clinical and genetic investigations are needed to fully understand the disease mechanism(s) of the entire entity of ABCA4-associated retinopathy, including childhood-onset and rapidly progressive retinal degeneration.

In conclusion, the present study underscores the broad and variable mutational spectrum of the largest cohort of STGD1 to date, including the reporting establishing of more than 50 novel likely pathogenic sequence variants. The high proportion (50%) of patients harbouring only missense variants is compatible with the relatively mild phenotype of the ProgStar cohort as a whole. There is a suggestion that geographic area is associated with the relatively mild phenotype of the ProgStar cohort as a whole. Likely pathogenic sequence variants. The high proportion (50%) of patients harbouring only missense variants is compatible with the relatively mild phenotype of the ProgStar cohort as a whole.

**Contributors**

HPNS and MM have full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HPNS and MM have full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HPNS and MM have full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HPNS and MM have full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HPNS and MM have full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Foundation, New York City, New York, USA; Ocular Albinism Research Fund (Clark Enterprises); unrestricted grant to the Wilmer Eye Institute from Research to Prevent Blindness; Baylor-Johns Hopkins Center for Mendelian Genetics (National Human Genome Research Institute, NHGRI/NIH; identification no: 1U5HG00654-02).}

**Competing interests**

KF is a paid consultant of Astellas Pharma, Kubota Pharmaceutical Holdings. DGB is a consultant for NightStarRx, AGTC, Shire, Ionis and Genentech. EZ is a member of the Data Monitoring and Safety Board/Committee of the following entities: ReNeuron Group/Ora, NightStarRx, RD-CURE Consortium and principal investigator in clinical trials sponsored by QLT, SHIRE and FFB at the University of Tuebingen, Institute for Ophthalmic Research. HPNS is a paid consultant of the following entities: Boehringer Ingelheim Pharma, Daiichi Sankyo, Gerson Lehrman Group; Guidepoint and Shire. HPNS is member of the Scientific Advisory Board of the Astellas Institute for Regenerative Medicine; Gensis Biologics; Vision Medicines and Intella Therapeutics. HPNS is member of the Data Monitoring and Safety Board/Committee of the following entities: Genentech/F. Hoffmann-La Roche and ReNeuron Group/Ora. These arrangements have been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. Johns Hopkins University and Bayer Pharma have an active research collaboration and option agreement. These arrangements have also been reviewed and approved by the University of Basel (Universitätsspital Basel, USB) in accordance with its conflict of interest policies. HPNS is principal investigator of grants at the USB sponsored by the following entity: Acucela; NightstaRx; QLT. Grants at USB are negotiated and administered by the institution (USB) which receives them on its proper account.

**Patient consent**

Not required.

**Ethics approval**

NHS National Research Ethics Service.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Open access**

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited, any changes made indicated, and the use and license their derivative works on different terms, provided the original work is ethic approval.

**Patient consent**

Not required.

**Ethics approval**

NHS National Research Ethics Service.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Open access**

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited, any changes made indicated, and the use and license their derivative works on different terms, provided the original work is. 2019;54:847–501.


