HGG Advances

ARTICLE

Targeted sequencing and *in vitro* splice assays shed light on *ABCA4*-associated retinopathies missing heritability

Zelia Corradi,[1,18,*] Mubeen Khan,[1,2] Rebekkah Hitti-Malin,[1] Ketan Mishra,[1] Laura Whelan,[3] Stéphanie S. Cornelis,[1] ABCA4-Study Group, Carel B. Hoyng,[4] Kati Kämpjärvi,[5] Caroline C.W. Klaver,[4,6,7,8] Petra Liskova,[9,10] Heidi Stöhr,[11] Bernhard H.F. Weber,[11,12] Sandro Banfi,[13] G. Jane Farrar,[3] Dror Sharon,[14] Jana Zernant,[15] Rando Allikmets,[15,16] Claire-Marie Dhaenens,[1,17] and Frans P.M. Cremers[1]

Summary

The *ABCA4* gene is the most frequently mutated Mendelian retinopathy-associated gene. Biallelic variants lead to a variety of phenotypes, however, for thousands of cases the underlying variants remain unknown. Here, we aim to shed further light on the missing heritability of *ABCA4*-associated retinopathy by analyzing a large cohort of macular dystrophy probands. A total of 858 probands were collected from 26 centers, of whom 722 carried no or one pathogenic *ABCA4* variant, while 136 cases carried two *ABCA4* alleles, one of which was a frequent mild variant, suggesting that deep-intronic variants (DIVs) or other *cis*-modifiers might have been missed. After single molecule molecular inversion probes (smMIPs)-based sequencing of the complete 128-kb *ABCA4* locus, the effect of putative splice variants was assessed *in vitro* by midigene splice assays in HEK293T cells. The breakpoints of copy number variants (CNVs) were determined by junction PCR and Sanger sequencing. *ABCA4* sequence analysis solved 207 of 520 (39.8%) naive or unsolved cases and 70 of 202 (34.7%) monoallelic cases, while additional causal variants were identified in 54 of 136 (39.7%) probands carrying two variants. Seven novel DIVs and six novel non-canonical splice site variants were detected in a total of 35 alleles and characterized, including the c.6283-321C>G variant leading to a complex splicing defect. Additionally, four novel CNVs were identified and characterized in five alleles. These results confirm that smMIPs-based sequencing of the complete *ABCA4* gene provides a cost-effective method to genetically solve retinopathy cases and that several rare structural and splice altering defects remain undiscovered in Stargardt disease cases.

Introduction

ABCA4 is the most frequently mutated Mendelian retinopathy-associated gene.[1] Biallelic variants in the gene lead to a variety of phenotypes, including autosomal recessive Stargardt disease (STGD1)[2] and cone-rod dystrophy.[3] The *ABCA4* gene encodes the transmembrane ATP-binding cassette transporter type A4 (ABCA4), which is located in the rims of photoreceptor outer segment discs[4] and functions as a retinoid transporter facilitating the clearance of potentially toxic derivates of the visual cycle.[5] ABCA4 dysfunction leads to the formation and accumulation of cytotoxic bisretinoid compounds (collectively called lipofuscin) leading to retinal pigment epithelium damage and photoreceptor cell death, ultimately causing progressive loss of vision.[6,7]

Clinical manifestation of the disease is characterized by a large phenotypic heterogeneity and variable age of onset.[8,9] In classic STGD1, loss of central vision starts around the second decade of life, but both early- and late-onset subtypes have been extensively described.[10,11,12,13,14] Disease variability depends on the specific combination of alleles found in each patient, where variants can be classified in different severity categories from mild to deleterious.[9,15] The identification of two pathogenic alleles segregating with the disease is essential to provide an accurate clinical prognosis and determine the eligibility of patients

[1]Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands; [2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; [3]The School of Genetics & Microbiology, Trinity College Dublin, Dublin, Ireland; [4]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, the Netherlands; [5]Blueprint Genetics, Espoo, Finland; [6]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; [7]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands; [8]Institute of Molecular & Clinical Ophthalmology, Basel, Switzerland; [9]Research Unit for Rare Diseases, Department of Paediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; [10]Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; [11]Institute of Human Genetics, University of Regensburg, Regensburg, Germany; [12]Institute of Clinical Human Genetics, University Hospital Regensburg, Regensburg, Germany; [13]Department of Precision Medicine, University of Campania "Luigi Vanvitelli," Naples and Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; [14]Department of Ophthalmology, Hadassah Medical Center, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; [15]Department of Ophthalmology, Columbia University, New York, NY, USA; [16]Department of Pathology & Cell Biology, Columbia University, New York, NY, USA; [17]University Lille, Inserm, CHU Lille, U1172 - LilNCog - Lille Neuroscience & Cognition, 59000 Lille, France
[18]Lead contact
*Correspondence: zelia.corradi@radboudumc.nl
https://doi.org/10.1016/j.xhgg.2023.100237.

for future gene- or mutation-specific therapies or trials.[8,16]

To date, more than 2,400 unique *ABCA4* variants have been published (LOVD: http://www.lovd.nl/ABCA4, accessed on 26.06.2023).[15] In particular, a high incidence of causative non-canonical splice site (NCSS) variants and deep-intronic variants (DIVs) has been observed in *ABCA4*.[8,15–21] Nevertheless, in many patients with phenotypes consistent with *ABCA4*-associated disease the underlying genetic causes are still unknown. In addition to frequent phenocopies, this missing heritability can be partially explained by the presence of splice altering or structural variants (SVs) that elude identification during diagnostic screening. Next to gene-panel sequencing, whole exome sequencing (WES) is preferred in a diagnostic setting because it entails lower sequencing and storage costs compared with whole genome sequencing. In contrast, most WES approaches lack the ability to identify DIVs and cannot define the boundaries of large SVs, when these are located in intronic regions of the gene. Moreover, *ABCA4*-associated retinopathies are characterized by a triad of clinical features (i.e., macular atrophy, lipofuscin deposits, and peripapillary sparing) that facilitate clinical identification, rendering *ABCA4* a suitable target for locus sequence analysis. Initially, the entire *ABCA4* locus was sequenced using RainDance microdroplet-PCR target enrichment,[21] while more recently single molecule molecular inversion probes (smMIPs)-based sequencing enabled the analysis of 1,054 STGD probands.[19]

In this study, we aim to shed further light on the missing heritability of macular dystrophy and report the results of smMIPs-based sequencing in 722 unsolved macular dystrophy cases as well as additional findings in 136 probands in whom two variants were already identified. Subsequent *in silico* assessment of potential splice altering variants led to the identification of 13 novel splice-altering variants, including seven causal DIVs and six causal NCSS variants that were characterized by *in vitro* splice assays. These results highlight the effectiveness of whole locus sequencing to genetically solve macular dystrophy cases and the importance of efficiently identifying and characterizing splice altering variants to shed further light on the missing heritability of macular dystrophy.

Materials and methods

Cohort composition

The study cohort consisted of 858 probands with phenotypes consistent with *ABCA4*-associated retinopathy (Table S1), which were collected in 26 international centers (Table S2). All samples were derived from probands for which STGD was part of the differential diagnosis. Among the 858 probands, no previous sequencing analysis had been performed in 355 cases, while 503 cases were previously screened using different methods (Table S2). Based on the status before smMIPs sequencing, two distinct patient groups could be determined. The first group comprised 722 unsolved probands who carried one (n = 202) or no (n = 520) *ABCA4* pathogenic alleles. The second patient group consisted of 136 cases with at least two variants in which one of the two identified alleles was a frequent, mild allele such as (p.(Asn1868Ile) and p.(Gly1961Glu)), or a benign allele such as p.[Gly863Ala,Gly863del]. The latter variant was considered a pathogenic allele in its own right until 2017 when Zernant et al.[14] showed that p.[Gly863Ala,Gly863del] was completely penetrant and pathogenic only when found in *cis* with p.(Asn1868Ile). These biallelic probands were included, since it is possible that an unidentified modifier or DIV is present in *cis* with these variants. Samples were collected according to the tenets of the Declaration of Helsinki and written informed consent was obtained for all patients participating in the study. DNA samples were quantified, diluted to 15 ng/μL and DNA integrity assessed using agarose gel electrophoresis.

smMIPs-based *ABCA4* sequence analysis and data processing

Sequencing of the samples was performed using an in-house smMIPs probe pool, as described previously.[19] Briefly, 3,866 smMIPs were designed to capture and sequence the entire *ABCA4* gene, in addition to 23.9 kb upstream and 14.9 kb downstream regions. Each probe has a capture region of 110 nt and the overall smMIPs pool was designed to cover every nucleotide position in *ABCA4* with two different smMIPs, thereby increasing sensitivity and minimizing dropout because of variants in the smMIPs arms. For SNPs known to be in the sequence of the annealing arms, different smMIPs were designed to avoid allelic dropout.

To sequence the 858 cases, five sequencing runs (runs 1–5) of ∼215 samples each, including 6 control samples, were performed on an Illumina NextSeq 500 platform using high-output kits (∼400 million reads). Run 3 was performed twice because of the low coverage obtained in the first sequencing run. A threshold of >100,000,000 reads per run and >150× average smMIPs coverage was applied for all runs. Variant mapping and annotation were performed using an in-house bioinformatics pipeline, as previously described,[22] on raw sequencing data files (FASTQ). Sequencing reads were aligned to the reference genome using the human genome build GRCh37/hg19. Variant annotation included gene component information and the predicted effect of the variants, as well as frequency information derived from several population databases.

Data analysis and variant selection

Data analysis was performed in two steps. Firstly, all known *ABCA4* variants were identified, i.e., variants that were previously reported in literature and splice-affecting variants that had been tested using *in vitro* splice assays or patient-derived photoreceptor progenitor cells.[16,17,19,23–29] During this step data in the *ABCA4* Leiden Open Variation Database (LOVD: www.lovd.nl/ABCA4)[15,18] was used for the identification of variants known in literature.

Novel causative variants were selected by applying filters on the annotated data. Variants were filtered based on their position (e.g., exon, canonical splice site sequences, NCSS). Variants were considered to be NCSS variants when found in nucleotide positions −14 to −3 upstream of exons, +3 to +6 nucleotides positions downstream of exons, or positioned within the first and last two nucleotides of an exon. Allele frequency (AF) of <0.005 filtering was performed both on the internal sequencing run frequency and population databases, such as the Genome Aggregation

Database (gnomAD v2.1.1; http://gnomad.broadinstitute.org/, average AF) and Genome of the Netherlands (GoNL; https://www.nlgenome.nl/). Variant effect was assessed using *in silico* pathogenicity tools, such as Sorting Intolerant From Tolerant (SIFT; https://sift.bii.a-star.edu.sg),[30] Polymorphism Phenotyping v2 (PolyPhen-2; http://genetics.bwh.harvard.edu/pph2/),[31] and Mutation Taster (https://www.mutationtaster.org/).[32] Additionally, variants were checked through the Franklin Genoox platform (https://franklin.genoox.com/: Accessed July 2022) for suggested American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) classification.

The second phase of analysis focused on identifying novel putative splice altering variants. In addition to the filter on AF (<0.005), the effect of these variants on splicing was assessed considering SpliceAI predictions.[33] SpliceAI was set to return the highest delta score in a window of 5,000 nt upstream and downstream of the variant position. Candidate variants were selected with a delta score $\geq$0.03 in either acceptor gain (AG), donor gain (DG), acceptor loss, or donor loss. To ensure data were of high quality, the region read coverage and the presence of the variant were checked in each sample BAM file using Alamut Visual 2.13 software. Additionally, the programs available on Alamut Visual 2.13 software, that is, SpliceSiteFinder-like,[34] MaxEntScan,[35] NNSPLICE[36] and GeneSplicer[37] were used as aid to visualize the position and predicted effect of the variants.

Midigene-based splice assay

The effects of 18 non-coding variants (12 DIVs and 6 NCSS variants) on splicing were assessed through *in vitro* splice assays. Nine previously described minigenes and midigene wild-type constructs containing *ABCA4* genomic sequences (between 1.25 and 8.5 kb), based on the wild-type clones BA1 through BA28, were used[29] (Figure S1A). Site-directed mutagenesis was performed to generate a mutant construct of each variant (Table S3B), followed by Gateway Cloning to obtain expression vectors. Additionally, two new minigenes, BA37 (hg19: exon 49, g.94,462,212–94,461,217) and BA38 (hg19: intron 49, g.94,460,003–94,459,176) were generated by directly inserting amplified patient genomic DNA in the expression vector pCI-*NEO-RHO* (Figure S1B and Table S3A) and subsequently sequenced to exclude the presence of additional SNPs in the constructs.

Wild-type and mutant constructs were transfected in parallel in Human Embryonic Kidney T (HEK293T) cells. HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin, and 1% sodium pyruvate at 37°C and 5% CO_2. Transfection was performed in duplicate on cells at 70% confluency using 600 ng plasmid and FuGENE HD reagent (Promega, Madison, WI), as specified in the manufacturer's protocol.

After 48 h, RNA was extracted from the cells using the Nucleospin RNA kit (Machery-Nagel, Düren, Germany) and cDNA was synthesized from 1,000 ng RNA through the iScript cDNA Synthesis kit (Bio-Rad, Hercules, CA). Reverse-transcription (RT)-PCR conditions were 94°C for 2 min, followed by 35 cycles of 30 s at 94°C, 30 s at 58°C and 5 min at 72°C, with a final extension step of 2 min at 72°C. For amplification, *ABCA4* exonic primers were used whenever possible. Alternatively, primers for the flanking exon 3 and 5 of *Rhodopsin* (*RHO*) present in the midigene system were utilized. *Actin beta (ACTB)* was selected as housekeeping gene control and exon 5 of *RHO* as a transfection control (Table S3C). Electrophoresis gel analysis was performed, followed by Sanger sequencing of the bands excised from gel (Figures S2–S13). Ratios between different RNA products were quantified by densitometry analysis using Fiji software after gel electrophoresis,[38] in RNA and protein notations only products that accounted for more than 15% of total product were reported. The cDNA, RNA and protein notation of all variants were obtained using the MANE transcript NM_000350.3.

Copy number variant analysis

Detection of copy number variants (CNVs) was performed using an Excel script developed in-house based on the smMIP read depth, as previously described.[19] Since this method allows for the identification of large deletions and duplications but not of inversions and insertions, the term CNV is used hereafter instead of SV. CNV analysis was performed independently for each of the sequencing runs (Tables S4–S8), and CNVs were confirmed by PCR amplification using AmpliTaq Gold DNA polymerase (Thermofisher Scientific, Waltham, MA) followed by Sanger sequencing to identify the breakpoints of the rearrangement (Table S3D and Figure S14). Six samples containing known CNVs were included in all runs as positive controls.

Results

ABCA4 gene sequencing and variant identification

Sequencing of the complete *ABCA4* locus in 858 probands with phenotypes consistent with *ABCA4*-associated retinopathy revealed 287 unique (likely) pathogenic SNVs, 23 complex alleles (Tables S9 and S14) and 5 CNVs in 1,000 of 1,716 alleles. The 287 SNVs included 31 (10.8%) novel variants. Details on the *in silico* predictions for novel SNVs are provided in Table S10. Splice-altering variants represented 24% of all alleles, of which 8.8% were DIVs, and 15.2% NCSS and CSS variants (Figure 1A).

In the sub-cohort of 722 cases where one or no variant was known before sequencing, smMIPs analyses solved 277 (38.4%) of these cases (Figure S15A, Tables S11 and S12); the remaining 445 of 722 (61.6%) cases remained unsolved. The 277 newly solved cases include 70 formerly monoallelic patients for whom novel causal variants were identified. We identified splice altering variants in 41 of 70 (58.6%) of these cases and CNVs in 3 of 70 (4.3%) patients, in the remaining 26 of 70 (37.1%) one or more coding variants were identified. After sequencing, 180 of 722 (24.9%) cases were considered monoallelic, this sub-cohort included 119 previously designated monoallelic patients for whom no additional *ABCA4* variant was found and 61 naive or unsolved samples in which only one causal variant was identified. The remaining 265 of 722 (36.7%) cases consisted of 252 individuals in which no *ABCA4* variant was identified and 13 cases where the previously reported monoallelic variant was not confirmed by smMIPs. Samples were considered monoallelic when one variant categorized as class 3 (variant of uncertain significance), 4 (likely pathogenic) or 5 (pathogenic) following ACMG-AMP guidelines was reported,[39] including known variants with a high AF in the control population. Samples (n = 9) were classified as unsolved if

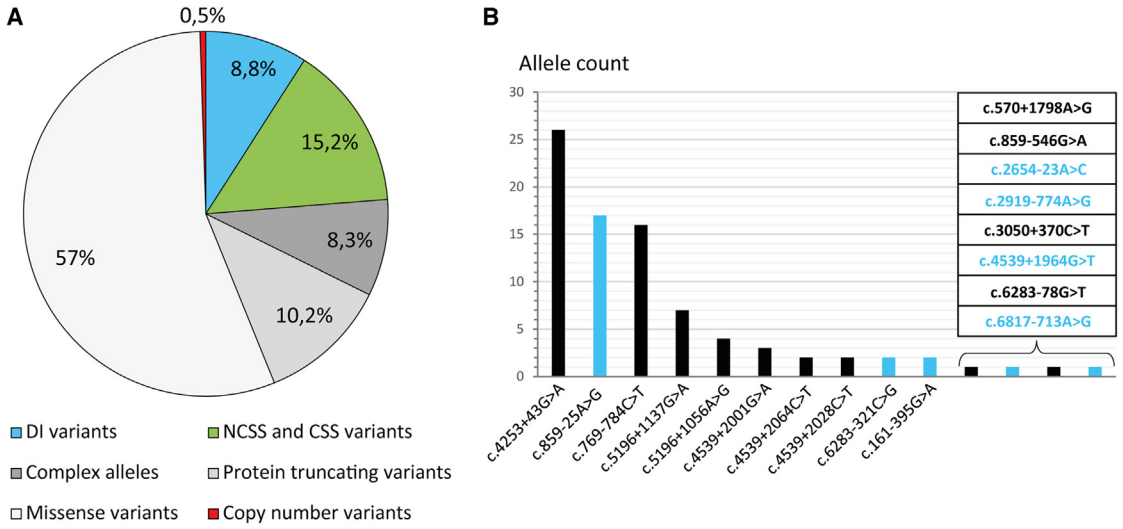

Figure 1. Distribution of *ABCA4* allele types in the sequenced cohort

(A) Distribution of all ABCA4-alleles present in biallelic and monoallelic cases based on the type of allele. Protein truncating variants refers to nonsense and frameshift variants in coding regions of the gene. In complex alleles only combinations of missense variants are counted. Complex alleles carrying deep intronic (DI), NCSS, and CSS variants are accounted in the respective categories.

(B) Number of alleles carrying known (in black) or novel (in blue) deep intronic variants. Eighteen unique DI variants were identified in 83 alleles. Except for variant c.859-25A>G, novel DI variants were found in one or two probands only.

two mild variants or one mild and one moderately severe variant were identified, as these combinations are not expected to lead to *ABCA4*-associated retinopathy on their own. As an exception, four cases carrying c.5882G>A in homozygosity were considered solved, as similar instances of patients carrying only this variant had been previously reported in other studies.[40]

In the sub-cohort of 136 biallelic cases additional causal variants were identified in 54 (39.7%), while there were no additional finding in 77 (56.6%) of cases. In the remaining five cases (3.7%), one or more of the previously reported variants were not confirmed. The most frequent previously not reported variants were c.5603A>T (n = 16) and c.769-784C>T (n = 10); the latter was always found in *cis* with the c.5882G>A variant, confirming the previously described complex allele[40] (Figure S15B, Tables S11 and S12). Apart from c.769-784C>T, splice altering variants were identified in 10 cases, including four samples carrying the novel NCSS variant c.4667+5G>T.

The most common pathogenic *ABCA4* alleles in the complete cohort were c.5882G>A (n = 150; 15%) and c.5603A>T (n = 122; 12.2%), followed by the complex alleles c.[2588G>C;5603A>T] and c.[5461-10T>C;5603A>T] with n = 37 (3.7%) and n = 31 (3.1%), respectively. Among non-coding variants, the most frequent were c.4253+43G>A (n = 26, 2.6%), the recently identified[41] founder mutation c.859-25A>G (n = 17, 1.7%) (Table S9 and Figure 1B) and c.769-784C>T (n = 16, 1.6%), which was found in *cis* with c.5882G>A or c.5603A>T.

Novel splice-altering variants

Data analysis based on *in silico* prediction tools identified 12 putative DIVs and 6 NCSS variants predicted to alter splicing (Table 1). All variants were selected based on a SpliceAI delta score (DS) of ≥0.03, with predictions ranging from 0.03 to 0.96 (Table S13). Of the 12 selected DIVs, 8 showed splicing alterations (Figures 2 and S16). The most frequent novel variant in the cohort (c.859-25A>G) was discovered, in parallel with this study, to be a novel founder branchpoint mutation in the Palestinian population. As the results concerning this variant have already been published,[41] this will not be discussed in more detail below. A second branchpoint variant (c.2654-23A>C) was identified in a single proband. Four variants (c.859-556A>G, c.2919-177G>T, c.6005+717G>T, c.6283-93C>G) did not show any effect (Figure S16). The highest SpliceAI DS predictions for these variants were DG = 0.16, AG = 0.03, AG = 0.1, and AG = 0.1, respectively. All six tested NCSS variants, which had DS predictions ranging from 0.24 to 0.76, showed splice altering effects *in vitro* (Figure 3).

All tested variants were classified into one of four severity classes based on the percentage of aberrant splicing observed: severe (≤20% wild-type RNA), moderately severe (>20% and ≤40% wild-type RNA), mild (>40% and ≤80% wild-type RNA), and benign (>80% wild-type RNA). Three DIVs were classified as moderately severe, one as mild and two as benign (Table 1). Of the NCSS variants, three were classified as severe, one as moderately severe, and two as mild (Table S14).

The moderately severe variant c.4539+1964G>T is situated in a hotspot region of intron 30, where four additional variants have already been reported.[16,23,42] All variants reported in this region share the same cryptic splice acceptor site at position g.94.493.110 (hg19) while the donor site utilized varies depending on the position of the causative

Table 1. Deep intronic (DI) and NCSS variants that were tested *in vitro* with midigene splice assay

cDNA variant (NM_000350.3)	Nr. of alleles	Type	% WT RNA	Severity[a]	RNA variant (NM_000350.3)	Protein variant (NM_000350.3)
c.161-395G>A	2	DI	36	moderate	r.[161_302del,=,160_161ins161-484_161–394]	p.[Cys54Serfs*14,=,Cys54Cysfs*75]
c.859-556A>G	1	DI	100	benign	r. =	p. =
c.859-25A>G[b]	17	DI	0	severe	r.[859_1099del,858_859ins859-685_859,858_859ins859-138 _859]	p.[Phe287Hisfs*7,Phe287Tyrfs*33,Phe287Leufs*3]
c.2654-23A>C	1	DI	6	severe	r.2653_2654ins2654-40_2654-1	p.(Gly885Alafs*48)
c.2743G>A	1	NCSS	2	severe	r.[2654_ 2743del,2743_2744ins2743+1_2743+152)	p.[Gly885_His914,Asp915Glyfs*35]
c.2919-774A>G	1	DI	33	moderate	r.[2918_2919ins2919-957_ 2919–779,=]	p.[Leu973Phefs*3,=]
c.2919-177G>T	2	DI	100	benign	r. =	p. =
c.3522+524A>G	1	DI	89	benign	r.[=, 3522_3523ins3522+369_3522+519]	p.[=,Glu1174_Gly1175ins*14]
c.4254-5T>A	2	NCSS	2	severe	r.[4254_4352del,4254_4539del,4254_4466del]	p.[Ser1418_Pro1451delinsArg,Ser1418Argfs*13,Ser1418_Cys1488del]
c.4539+1964G>T	1	DI	22	moderate	r.[4539_4540ins4539+1891_4539+1962,=]	p.[Gln1513_Arg 1514ins(24),=]
c.4667+5G>T	4	NCSS	16	severe	r.[4635_4667del,=]	p.[Ser1545_Gln1555del,=]
c.4773G>T	1	NCSS	53	mild	r.[=,4668_ 4773del]	p.[=,Tyr1557Alafs*18]
c.6005+717G>T	1	DI	100	benign	r. =	p. =
c.6283-321C>G	2	DI	90	benign[c]	r.[=,6282_6283ins6382+173_6283–322,6282_6283ins6382+1_6283–322]	p.[=,Leu2094_Asp2095ins*58,Leu2094_Asp2095ins*117]
c.6283-93C>G	1	DI	100	benign	r. =	p. =
c.6387–15_6387-12del	1	NCSS	36	moderate	r.[6386_6387ins 6386+1_6387–1,=]	p.[Ser2129Argfs*30,=]
c.6816G>A	1	NCSS	67	mild	r.[=,6730_ 6816del]	p.[=,Val2244_Gln2272del]
c.6817-713A>G	1	DI	59	mild	r.[=,6730_6816ins6817-835_6817–714]	p.[=,Gln2272_Asp2273ins*11]

[a]Severity was assigned based on wild-type (WT) RNA levels, considering the following intervals: benign (>80% WT RNA), mild (>40%, ≤80% WT RNA), moderate (>20%, ≤40% WT RNA), severe (≤20% WT RNA).
[b]Corradi et al.[41]
[c]All mutant products together represent 10% of all RNA products, rendering this, according to our classification, a benign variant. But it is possible that our quantification is not accurate enough or that the aberrant RNAs are underrepresented in this artificial assay system as the much smaller WT fragment is amplified more effectively.

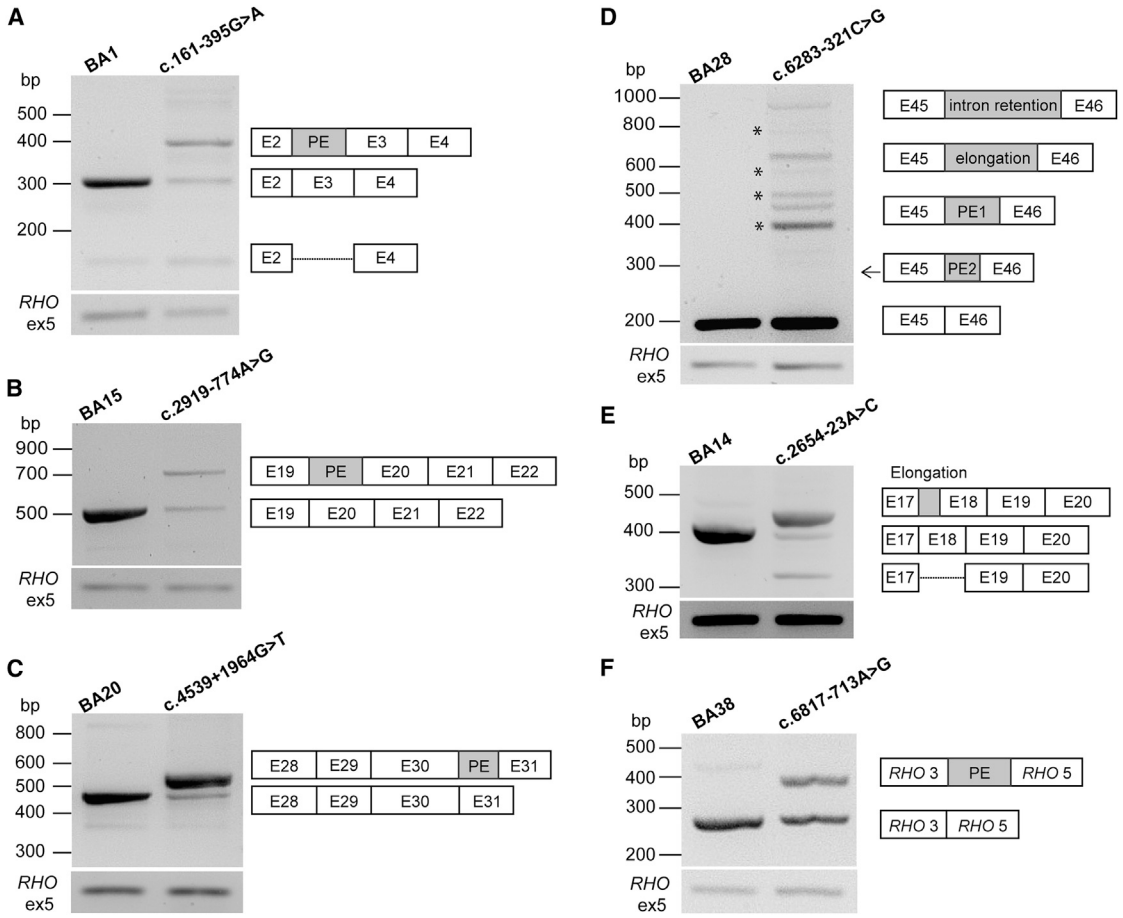

Figure 2. ***In vitro*** **splice assay results of novel causative deep intronic variants**
(A–F) In the transcript schematics, exons appear as white rectangles, pseudoexons and intron retentions as gray rectangles. Heteroduplex bands are tagged by an asterisk. Black arrows are used to signal faint bands for which Sanger sequencing validation was possible. Horizontal dotted lines are used to represent exon skipping.

variant. Variant c.4539+1964G>T leads to a pseudo-exon (PE) inclusion of 72 nt, causing an in-frame insertion of 24 amino acids. Interestingly, this was also the shortest PE inclusion identified in our cohort. Two DIVs (c.3522+524A>G and c.6283-321C>G) showed the expected splice defect, but aberrant transcripts were present in less than 20% of the cDNA products, leading to the tentative classification of these variants as benign (Table S14). Variant c.6283-321C>G showed a complex splicing defect with two PE inclusions and partial or complete intron inclusion (Figure 2). PE1 (258 nt) was the most abundant aberrant product, followed by the exon elongation, complete intron retention and PE2 (113 nt). PE1 corresponded with the highest predicted effect by SpliceAI, while PE2 was predicted as the fourth highest (Table S13). Several heteroduplex artifacts were observed after RT-PCR, which made result interpretation challenging. For this reason, to eliminate potential heteroduplex artifacts and improve the clarity of the results, additional RT-PCR experiments were performed using a lower number of cycles (33 cycles) (Figure S17). This approach allowed a clearer view of the main effects but also caused the loss of validated aberrant splicing bands (PE2). Considering this, semi-quantification was performed on the standard assay RT-PCR without considering the bands identified as heteroduplex, which means that the reported 10% aberrant splicing is likely to be an under-estimation of the total effect of the variant. The wild-type fragment observed for c.6283-321C>G (199 nt) is much smaller than all aberrant products (between 312 and 951 nt), so the larger mutant fragments will be vastly under-represented in this RT-PCR assay. We, therefore, consider that c.6283-321C>G is a pathogenic variant, but we cannot assess its severity.

Four of the NCSS variants showed one main aberrant product. Variants c.4667+5G>T, c.4773G>T and c.6816G>A led to single exon skipping, and the four-nucleotide deletion c.6387–15_6387-12del caused complete intron 45 retention. Conversely, the remaining variants exhibited more complex splicing defects. Variant c.2743G>A caused exon 18 skipping in 79% of the product and a 152-nt elongation in the remaining 19% (Figure 3) (Table S14). Variant c.4254-5T>A, affecting the splice acceptor site of exon 29, caused multiple splicing abnormalities that accounted for 98% of the total transcript. The main aberrant products were exon 29 skipping

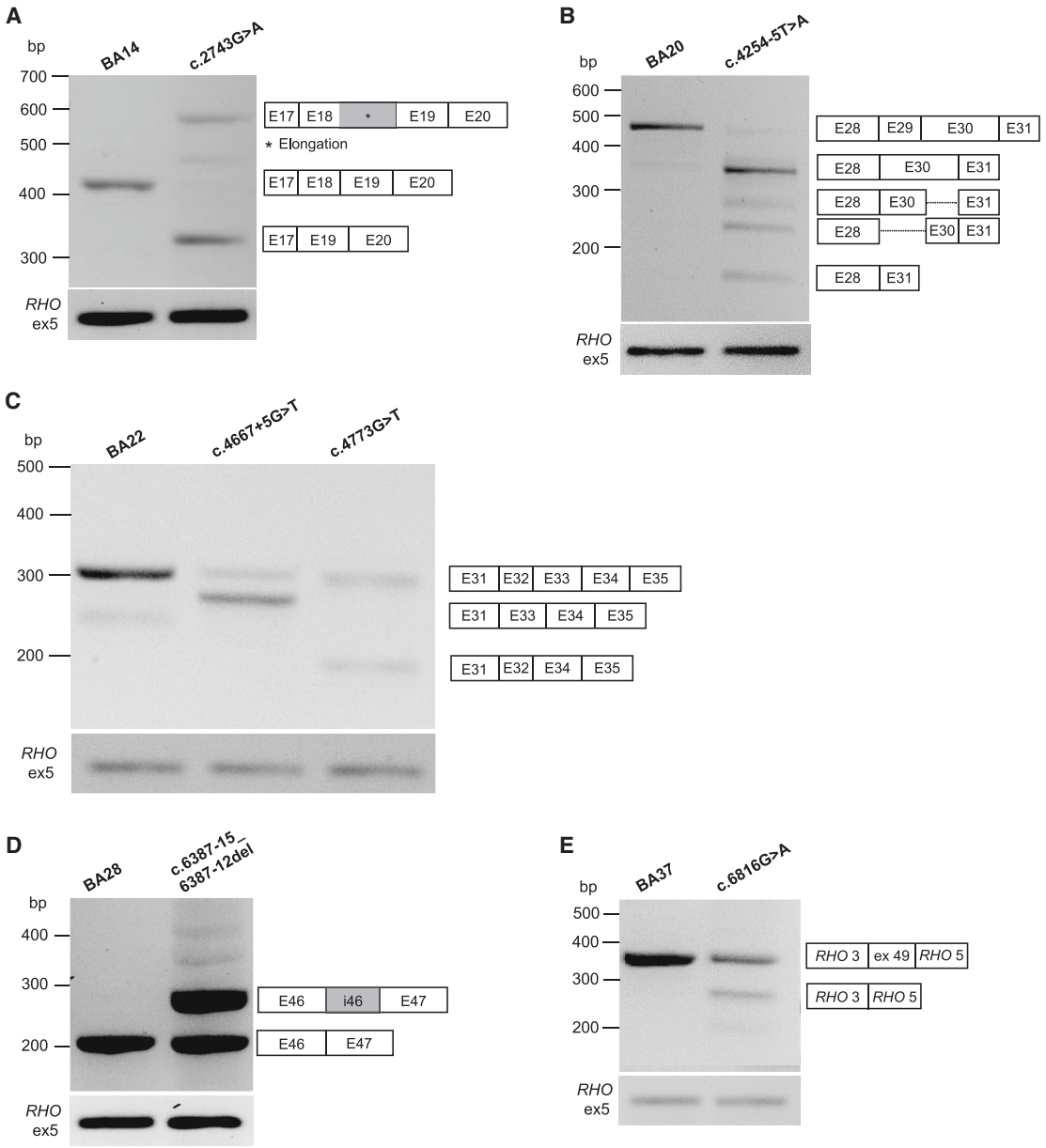

Figure 3. *In vitro* splice assay results of novel causative NCSS variants

(A–E) In the transcript schematics, exons appear as white rectangles, partial or complete intron retention as gray rectangles. In exon elongations, vertical dashed lines represent a direct continuation of the sequence from the adjacent exon. Horizontal dotted lines are used to represent partial exon skipping.

(44% of transcript) and exons 29–30 skipping (25% of the transcript). Exon 29 skipping was also observed in combination with partial exon 30 skipping, where either the first 114 nt (19% of transcript) or the last 73 nt (10% of transcript) of the exon were absent. Finally, the novel NCSS variant c.4667+5G>T was identified in four probands from Poland and was always inherited as part of a complex allele, c.[2549A>G;4667+5G>T;5882G>A], suggesting that this a founder allele (Table S15).

CNVs

In four probands, novel heterozygous CNVs were identified, consisting of three deletions and a tandem duplication (Figure 4 and S14). All three deletions span exonic regions and encompass 1,38 to 6.99 kb. c.442+1194_571-1399del spans exon 5 (sample ID: DNA11-05475, Table S5), c.2918+552_3329-601del encompasses exons 20 to 22 (sample ID: 071836, Table S5) and c.4667+507_4849-94del encompasses exons 33 and 34 (sample ID: 071730, Table S5). In one sample (ID: 073202, Table S8), a large tandem duplication of over 8.5 kb, encompassing exon 7, was identified (c.768+6839_858+66dup). Additionally, the known small deletion c.699_768+341del was identified in a patient of Spanish origin (ID: 072826) (Table S7). This CNV was frequently found in the Spanish population.[19,43] CNV breakpoints

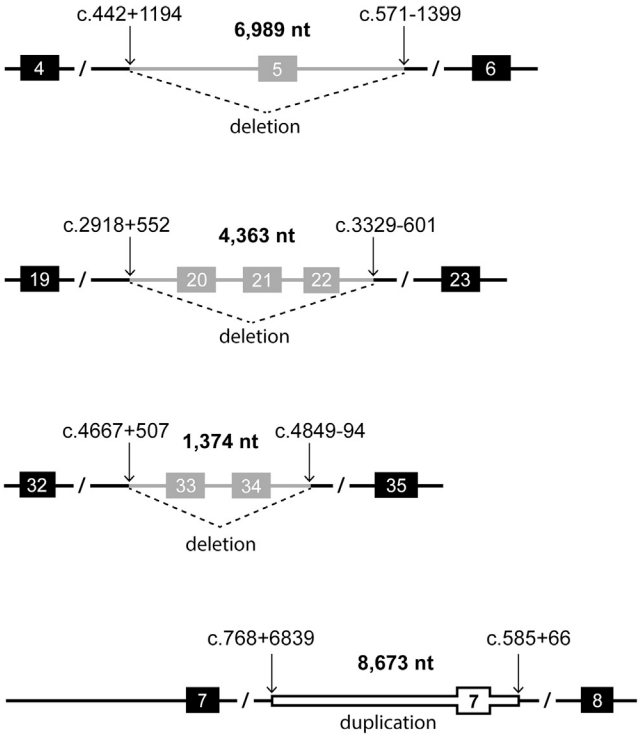

Figure 4. Novel CNVs
Novel CNVs consist of three deletions ranging in length from 1.4 kb to 7.0 kb and duplication of ~8.6 kb. In light gray are highlighted deleted regions, while the duplicated region is depicted as a bold, white-filled section.

were determined using genomic PCR and Sanger sequencing (Figure S14). For the duplication, additional PCR amplifications were performed to determine the orientation of the duplicated segment, which is head-to-tail. In all the CNVs, except for deletion c.2918+552_3329-601del, microhomology was observed at the breakpoints (2 or 3 nt). Consequently, the exact breakpoints could not be exactly identified and the notations are reported following the 3′ rule of HGVS nomenclature (HGVS: varnomen.hgvs.org).

Discussion

Sequencing of the *ABCA4* locus in 858 probands with phenotypes consistent with *ABCA4*-associated retinopathy using smMIPs lead to the identification of two or more pathogenic variants in 38.4% of a genetically unsolved cohort of 722 probands, which is in line with the solve rates observed in similar cohorts,[19,22] and the identification of additional findings in 39.7% of pre-screened cases with two known causative variants. *In silico* and *in vitro* analysis of non-coding sequencing data led to the identification of six novel NCSS and seven novel DIVs causing splicing defects. In addition, four novel heterozygous CNVs were identified. Overall, splice altering variants were identified in 240 alleles (24% of all identified alleles), of which 41 (17% of splice altering alleles) were novel findings in previously unsolved monoallelic cases. This highlights the relevance of screening intronic regions to identify causative variants in *ABCA4* both for naive cases and for unsolved samples which could potentially carry known or novel splice-altering variants.

The use of SpliceAI as the primary tool for the selection of putative splice altering variants resulted in only a few false positives (4/18; 22%), despite the arbitrarily conservative threshold set for the initial analysis (DS $\geq$ 0.03). In previous studies that used alternative *in silico* methods of variant assessment (e.g., Alamut prediction tools) the percentage of reported false positives was considerably higher (45/58; 77.6%).[19] Moreover the use of SpliceAI as primary prediction tool allowed the identification of two branchpoint variants (c.859-25A>G and c.2654-23A>C) that were not flagged by other prediction algorithms. The percentage of correctly predicted variants using SpliceAI is likely to become even higher after re-assessment of the ideal DS threshold to use for variant selection, as all DIVs leading to a confirmed effect in this study had a DS $\geq$ 0.20. Having a reliable cut-off for variant selection is important to ensure the identification of all causative variants while maintaining a low number of false positives, allowing for cost-effective studies. Nevertheless, careful consideration needs to be taken in the selection of the DS cutoff, as more stringent filtering carries the risk of missing potential variants that lead to milder effects. Splice-altering variants leading to a low percentage of aberrant splicing (<20%) are classified as benign and are not considered as a causative allele on their own. Nevertheless, the effect of such variants, of which we identified three in our cohort (c.2931G>A, c.3522+524A>G, c.6283-321C>G), cannot be completely discarded, as they could contribute to pathogenicity as *cis* modifiers in complex alleles. In this cohort, 21 different complex alleles were identified in a total of 126 alleles, of which 36 included splice altering variants, including the novel c.[2549A>G; 4667+5G>T; 5882G>A] allele (Table S15). Finally, the deep learning SpliceAI tool was not trained on retinal tissue, suggesting that retina-specific defects might be more difficult to predict and identify. In contrast, the *in vitro* splice assay based on HEK293T cells remains an artificial system that does not consider the potential role of retina-specific factors in the *in vivo* effect of splice-altering variants. It is thus possible that the effect of benign variants with low levels of aberrant splicing is currently being underestimated. Specifically, c.6283-321C>G showed a complex splicing defect comprising of three separate aberrant splicing events. Interpreting the variant effect proved challenging, both caused by the complexity of the defect and to the presence of several unclear artifacts in the midigene assay protocol, further compounding the underestimation of the full effect of the variant.

Two additional variants (c.2743G>A and c.4254-5T>A) showed complex splicing defects, categorized respectively as null and severe alleles. The c.4254-5T>A variant resulted in exon 29 skipping as the main effect, but showed an

additional array of combined skipping between exon 29 and the complete or partial skipping of exon 30. Exon and intron 30 of *ABCA4* are known hotspots for splicing defects,[16,17,21,28] facilitated by the presence of multiple cryptic splice sites of varying strength found in the exonic and intronic areas. In exon 30, partial skipping is likely influenced by the sequence AGGT at position c.4465_4468, which, together with suitable surrounding sequences, can function as either a cryptic splice acceptor or donor site. In particular, the acceptor site at position c.4466 has a predicted strength of 84.5 (SpliceSiteFinder-like, [0–100]) making it stronger than the canonical splice acceptor site itself (80.0 in SpliceSiteFinder-like). In contrast, the strength of the cryptic SDS is lower (67.7 in SpliceSiteFinder-like) than the canonical splice donor signal (76.5 in SpliceSiteFinder-like). Taken together, these scores support the result obtained in the splicing assay, where skipping of the first part of exon 30 was observed as the third most abundant transcript (19% of total transcript), while skipping of the second part of the exon only accounted for 10% of the total transcript. As with the variant c.6283-321C>G, the use of the HEK293T cells might not be providing the complete picture of the defect. While it has been demonstrated that this system is both reliable and cost effective,[29] further analysis of complex variants should be conducted using systems with a more accurate genomic and retinal context, such as patient-derived photoreceptor precursor cells or retinal organoids, to obtain a more precise view of the variant effect.

Most identified DIVs lead to a frameshift with the creation of a premature stop codon and consequently the production of a truncated protein, but in two cases predicting the repercussions of the variants for the ABCA4 protein was more challenging. The c.4539+1964G>T variant caused an in-frame PE insertion between exon 30 and 31. This could potentially affect protein function as the PE is located in the region encoding the extracellular domain 2 of ABCA4. Similarly, the effect of variant c.6817-713A>G is unclear due to its position in intron 49, at only one amino acid away from the wild-type sequence stop codon. As the PE itself contains a stop codon after the insertion of 10 amino acids, the overall effect of this PE inclusion is difficult to assess solely based on the performed splice assay. Further functional analysis of these two variants would shed light on the precise mechanism of pathogenicity.

Overall, the number of novel splice-altering variants identified and confirmed to alter splicing (n = 12) was lower than in previous studies (n = 22,[19]) and were mainly identified in one or two probands, with the striking exception of the Palestinian founder mutation c.859-25A>G, which was found in 17 alleles in 10 probands. These results suggest that, in non-White populations, frequent splice-altering variants remain to be identified and characterized, while in populations of predominant European descent, yet undiscovered DIVs are rare or ultra-rare occurrences.

Sequencing the entire *ABCA4* genomic locus in naive or unsolved samples (no variant identified in pre-screening) yielded a solve rate of 39.8% (207/520 samples), while in the subcohort of 202 unsolved monoallelic cases additional causal variants were identified in 34.7% (70/202) of samples. Taken together, 38.4% of the unsolved cohort could be explained by *ABCA4* variants after smMIPs sequencing. The remaining 445 probands remained unsolved, in 180 cases only one variant and in 265 patients no causal *ABCA4* variants were identified. In 119 of 202 of the monoallelic cases, no additional disease-associated variants were found after smMIPs sequencing. Considering the estimated carriership of *ABCA4* pathogenic variants in the general population to be approximately 14%–15%,[15] it is possible that part of the heterozygous variants identified in monoallelic cases are chance findings. Overall, it is likely that part of the missing heritability observed in the cohort could be explained by variants in other IRD-associated genes, as both the clinical heterogeneity and the presence of known phenocopies of STGD renders a correct clinical diagnosis challenging. Thus, sequencing of other known macular dystrophy-associated genes is likely to yield a genetic diagnosis for a portion of the samples for which two *ABCA4* variants were not identified in this study.[44,45] Additionally, it is possible that, despite the high coverage of the smMIPs sequencing runs (average nucleotide coverage 700×), some causal alleles were not identified. CNV analysis, based on coverage, does not allow the identification of inversions and insertions. Unsolved probands might also carry variants affecting regulatory elements present both in the open reading frame of the gene or in the untranslated regions. Furthermore, upstream and downstream regions of *ABCA4* were not screened for potential regulatory variants in this study.

The study cohort also included 136 samples in which two variants had been identified before smMIPs sequencing, mainly consisting of cases carrying the frequent mild alleles p.(Asn1868Ile), p.(Gly1961Glu), and the benign variant p.[Gly863Ala,Gly863del]. As the complete *ABCA4* locus was not covered in pre-screening, these samples were included to assess the possible presence of DIVs or other *cis*-modifiers that were previously missed. After smMIPs-based sequencing, splice altering variants were identified in 20 of 136 (14.7%) samples, including three novel variants (c.303-1G>T, 4667+5G>T and c.6816G>A). The DI variant c.769-784C>T was identified in eight samples carrying c.5882G>A and one sample carrying c.5603A>T, which are known to form frequent complex alleles with c.769-784C>T. While segregation analysis was possible only for the sample carrying c.[769-784C>T;5603A>T] as a complex allele, c.769-784C>T is well known to act as a *cis*-modifier of c.5882G>A.[40] Variant c.769-784C>T was also found in a proband carrying c.1928T>G and c.3874C>T. To our knowledge, neither variant has been previously reported to form a complex allele with c.769-784C>T. The most frequent novel finding was the missense variant p.(Asn1868Ile) which was present in 24 samples. This can be explained by the fact that most of these cases had undergone pre-screening before the causality of p.(Asn1868Ile) as an hypomorphic variant was recognized.[14,20,46] Notably,

p.(Asn1868Ile) was found as part of the complex allele c.[2588G>C;5603A>T] in 15 of 16 of the cases where previously only variant p.[Gly863Ala,Gly863del] was reported, which confirms that the presence of both variants is required to have a fully penetrant pathogenic allele.[14]

In conclusion, sequencing of the *ABCA4* locus in 858 probands with phenotypes consistent with *ABCA4*-associated retinopathy solved 47.1% of cases. The identification of 12 novel splice-altering variants, three of which cause complex splicing defects, and four new CNVs, further expands the list of disease-causing *ABCA4* variants.

Data and code availability

Data are available upon request. The RNA and protein products of all pathogenic variants found as a result of smMIPs sequencing have been deposited into the LOVD for *ABCA4* (LOVD: http://www.lovd.nl/ABCA4). All other smMIPs sequencing data are subject to controlled access because of research participant privacy. These data may become available upon a data transfer agreement approved by the local ethics committee and can be obtained upon reasonable request.

Consortia

ABCA4-Study Group: A.AlTalbishi, C. Ayuso, E. Banin, T. Ben-Yosef, H. Bolz, L.I. van den Born, A. Dockery, S. Downes, A. Fakin, M.B. Gorin, E. Heon, C.F. Inglehearn, M. Karali, M. Oldak, A.S. Plomp, J. Sajovic, J.P. Szaflik, A.A.H.J. Thiadens, A. Tracewska, M. Vajter, J.B.G.M. Verheij.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2023.100237.

Acknowledgments

We thank Saskia van der Velde-Visser for technical assistance.

This work was supported by the Retina UK, grant no. GR591 (to F.P.M.C.), a Horizon 2020, Marie Sklodowska-Curie Innovative Training Network entitled European Training Network to Diagnose, Understand and Treat Stargardt Disease, a Frequent Inherited Blinding Disorder-StarT (813490) (to F.P.M.C.), the Foundation Fighting Blindness USA, grant no. PPA-0517-0717-RAD (to F.P.M.C.), the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, and the Stichting tot Verbetering van het Lot der Blinden (to F.P.M.C.), and by the Landelijke Stichting voor Blinden en Slechtzienden, the Macula Degeneratie fonds and the Stichting Blinden-Penning that contributed through Uitzicht 2016–12 (to F.P.M.C.). This work was also supported by the Pro Retina Foundation Germany, the Stichting Blindenhulp, the Stichting ter Verbetering van het Lot der Blinden, the Gelderse Blinden Stichting, the Stichting voor Ooglijders, and the Oogfonds that contributed through Uitzicht 2020–17 (to F.P.M.C.). This work was also supported by EJPRD19-234 (Solve-RET) (to F.P.M.C. and P.L.) and by the GA UK n.321522 (to P.L. and M.V.). This work was also supported by the Israel Science Foundation (grant number 1778/20 to D.S. and E.B.) within the Israel Precision Medicine Partnership program. This research was funded by Fighting Blindness Ireland (FB18CRE, FB20DOC). This work was also supported, in part, by the NIH/NEI grants R01EY028203, R01EY028954, R01EY029315, P30EY019007, Foundation Fighting Blindness award PPA-1218-0751-COLU, and the Unrestricted funds from the Research to Prevent Blindness (RPB) to the Department of Ophthalmology, Columbia University, New York, NY, USA.

Declaration of interests

The authors declare no competing interests.

Received: June 5, 2023
Accepted: September 8, 2023

Web resources

LOVD: http://www.lovd.nl/ABCA4

Franklin Genoox: https://franklin.genoox.com/

Genome Aggregation Database (gnomAD): http://gnomad.broadinstitute.org/

Genome of the Netherlands (GoNL): https://www.nlgenome.nl/

Sorting Intolerant From Tolerant (SIFT): https://sift.bii.a-star.edu.sg.

Polymorphism Phenotyping v2 (PolyPhen-2): http://genetics.bwh.harvard.edu/pph2/

Mutation Taster: https://www.mutationtaster.org/

References

1. Blacharski, P.A., and Newsome, D.A. (1988). Bilateral macular holes after Nd-Yag laser posterior capsulotomy. Am. J. Ophthalmol. *105*, 417–418.
2. 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.
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. Illing, M., Molday, L.L., and Molday, R.S. (1997). The 220-kDa rim protein of retinal rod outer segments is a member of the ABC transporter superfamily. J. Biol. Chem. *272*, 10303–10310.
5. Molday, R.S., Zhong, M., and Quazi, F. (2009). The role of the photoreceptor ABC transporter ABCA4 in lipid transport and Stargardt macular degeneration. Biochim. Biophys. Acta *1791*, 573–583.
6. Sparrow, J.R., and Boulton, M. (2005). "RPE lipofuscin and its role in retinal pathobiology." Exp. Eye Res. *80*, 595–606.
7. Sparrow, J.R., Wu, Y., Kim, C.Y., and Zhou, J. (2010). Phospholipid meets all-trans-retinal: the making of RPE bisretinoids. J. Lipid Res. *51*, 247–261.

8. Cremers, F.P.M., Lee, W., Collin, R.W.J., and Allikmets, R. (2020). Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. Prog. Retin. Eye Res. *79*.
9. Lee, W., Zernant, J., Su, P.Y., Nagasaki, T., Tsang, S.H., and Allikmets, R. (2022). "A genotype-phenotype correlation matrix for ABCA4 disease based on long-term prognostic outcomes." JCI Insight *7*.
10. Fujinami, K., Zernant, J., Chana, R.K., Wright, G.A., Tsunoda, K., Ozawa, Y., Tsubota, K., Robson, A.G., Holder, G.E., Allikmets, R., et al. (2015). "Clinical and Molecular Characteristics of Childhood-Onset Stargardt Disease." Ophthalmology *122*, 326–334.
11. Runhart, E.H., Valkenburg, D., Cornelis, S.S., Khan, M., Sangermano, R., Albert, S., Bax, N.M., Astuti, G.D.N., Gilissen, C., Pott, J.W.R., et al. (2019). Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. Invest. Ophthalmol. Vis. Sci. *60*, 4249–4256.
12. Tanaka, K., Lee, W., Zernant, J., Schuerch, K., Ciccone, L., Tsang, S.H., Sparrow, J.R., and Allikmets, R. (2018). The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. Ophthalmology *125*, 89–99.
13. Westeneng-van Haaften, S.C., Boon, C.J.F., Cremers, F.P.M., Hoefsloot, L.H., den Hollander, A.I., Hoyng, C.B., and Hoyng, C.B. (2012). "Clinical and genetic characteristics of late-onset Stargardt's disease." Ophthalmology *119*, 1199–1210.
14. Zernant, J., Lee, W., Collison, F.T., Fishman, G.A., Sergeev, Y.V., Schuerch, K., Sparrow, J.R., Tsang, S.H., and Allikmets, R. (2017). "Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration." J. Med. Genet. *54*, 404–412.
15. Cornelis, S.S., Runhart, E.H., Bauwens, M., Corradi, Z., De Baere, E., Roosing, S., Haer-Wigman, L., Dhaenens, C.M., Vulto-van Silfhout, A.T., and Cremers, F.P.M. (2022). Personalized genetic counseling for Stargardt disease: Offspring risk estimates based on variant severity. Am. J. Hum. Genet. *109*, 498–507.
16. Garanto, A., Duijkers, L., Tomkiewicz, T.Z., and Collin, R.W.J. (2019). "Antisense Oligonucleotide Screening to Optimize the Rescue of the Splicing Defect Caused by the Recurrent Deep-Intronic ABCA4 Variant c.4539+2001G>A in Stargardt Disease." Genes *10*, 452.
17. 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.
18. Cornelis, S.S., Bax, N.M., Zernant, J., Allikmets, R., Fritsche, L.G., den Dunnen, J.T., Ajmal, M., Hoyng, C.B., and Cremers, F.P.M. (2017). Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. Hum. Mutat. *38*, 400–408.
19. Khan, M., Cornelis, S.S., Pozo-Valero, M.D., Whelan, L., Runhart, E.H., Mishra, K., Bults, F., AlSwaiti, Y., AlTalbishi, A., De Baere, E., et al. (2020). "Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics." Genet. Med. *22*, 1235–1246.
20. Schulz, H.L., Grassmann, F., Kellner, U., Spital, G., Rüther, K., Jägle, H., Hufendiek, K., Rating, P., Huchzermeyer, C., Baier, M.J., Weber, B.H.F., and Stöhr, H. (2017). Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. Invest. Ophthalmol. Vis. Sci. *58*, 394–403.
21. Zernant, J., Xie, Y.A., Ayuso, C., Riveiro-Alvarez, R., Lopez-Martinez, M.A., Simonelli, F., Testa, F., Gorin, M.B., Strom, S.P., Bertelsen, M., et al. (2014). "Analysis of the ABCA4 genomic locus in Stargardt disease." Hum. Mol. Genet. *23*, 6797–6806.
22. Khan, M., Cornelis, S.S., Khan, M.I., Elmelik, D., Manders, E., Bakker, S., Derks, R., Neveling, K., van DeVorst, M., Gilissen, C., et al. (2019). "Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease." Hum. Mutat. *40*, 1749–1759.
23. Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoyng, C.B., Zernant, J., Lee, W., Allikmets, R., Collin, R.W.J., and Cremers, F.P.M. (2018). "Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease." Am. J. Hum. Genet. *102*, 517–527.
24. Bauwens, M., De Zaeytijd, J., Weisschuh, N., Kohl, S., Meire, F., Dahan, K., Depasse, F., De Jaegere, S., De Ravel, T., De Rademaeker, M., et al. (2015). "An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients." Hum. Mutat. *36*, 39–42.
25. Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., De Zaeytijd, J., Khan, M., Sadler, F., Balikova, I., Van Cauwenbergh, C., et al. (2019). ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. Genet. Med. *21*, 1761–1771.
26. Fadaie, Z., Khan, M., Del Pozo-Valero, M., Cornelis, S.S., Ayuso, C., Cremers, F.P.M., Roosing, S., Allikmets, R., Bauwens, M., Ghofrani, M., et al. (2019). "Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4." Hum. Mutat. *40*, 2365–2376.
27. Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A., Collin, R.W.J., Goercharn-Ramlal, A.S.A., den Engelsman-van Dijk, A.H.A., Rohrschneider, K., et al. (2016). Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T–>C Mutation in Stargardt Disease. Ophthalmology *123*, 1375–1385.
28. Sangermano, R., Garanto, A., Khan, M., Runhart, E.H., Bauwens, M., Bax, N.M., van den Born, L.I., Khan, M.I., Cornelis, S.S., Verheij, J.B.G.M., et al. (2019). "Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides." Genet. Med. *21*, 1751–1760.
29. Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A., Elmelik, D., Qamar, R., Lugtenberg, D., van den Born, L.I., et al. (2018). "ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease." Genome Res. *28*, 100–110.
30. Kumar, P., Henikoff, S., and Ng, P.C. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat. Protoc. *4*, 1073–1081.
31. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
32. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods *11*, 361–362.

33. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). "Predicting Splicing from Primary Sequence with Deep Learning." Cell *176*, 535–548.e24.
34. Leman, R., Gaildrat, P., Le Gac, G., Ka, C., Fichou, Y., Audrezet, M.P., Caux-Moncoutier, V., Caputo, S.M., Boutry-Kryza, N., Léone, M., et al. (2018). Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. Nucleic Acids Res. *46*, 7913–7923.
35. Yeo, G., and Burge, C.B. (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." J. Comput. Biol. *11*, 377–394.
36. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). "Improved splice site detection in Genie." J. Comput. Biol. *4*, 311–323.
37. Pertea, M., Lin, X., and Salzberg, S.L. (2001). GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. *29*, 1185–1190.
38. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.
39. Cornelis, S.S., Bauwens, M., Haer-Wigman, L., De. Bruyne, M., Pantrangi, M., Baere, D.E., Hufnagel, R.B., Claire-Marie, D., and Cremers, F.P.M. (2023). Compendium of clinical variant classification for 2,247 unique ABCA4 variants to improve genetic medicine access for Stargardt Disease. Preprint at medRxiv.
40. Lee, W., Zernant, J., Nagasaki, T., Molday, L.L., Su, P.Y., Fishman, G.A., Tsang, S.H., Molday, R.S., and Allikmets, R. (2021). "Cis-acting modifiers in the ABCA4 locus contribute to the penetrance of the major disease-causing variant in Stargardt disease." Hum. Mol. Genet. *30*, 1293–1304.
41. Corradi, Z., Salameh, M., Khan, M., Héon, E., Mishra, K., Hitti-Malin, R.J., AlSwaiti, Y., Aslanian, A., Banin, E., Brooks, B.P., et al. (2022). ABCA4 c.859-25A>G, a Frequent Palestinian Founder Mutation Affecting the Intron 7 Branchpoint, Is Associated With Early-Onset Stargardt Disease. Invest. Ophthalmol. Vis. Sci. *63*, 20.
42. Tomkiewicz, T.Z., Suarez-Herrera, N., Cremers, F.P.M., Collin, R.W.J., and Garanto, A. (2021). Antisense oligonucleotide-based rescue of aberrant splicing defects caused by 15 pathogenic variants in ABCA4. Int. J. Mol. Sci. *22*, 4621.
43. Del Pozo-Valero, M., Riveiro-Alvarez, R., Blanco-Kelly, F., Aguirre-Lamban, J., Martin-Merida, I., Iancu, I.F., Swafiri, S., Lorda-Sanchez, I., Rodriguez-Pinilla, E., Trujillo-Tiebas, M.J., et al. (2020). Genotype-Phenotype Correlations in a Spanish Cohort of 506 Families With Biallelic ABCA4 Pathogenic Variants. Am. J. Ophthalmol. *219*, 195–204.
44. Hitti-Malin, R.J., Dhaenens, C.M., Panneman, D.M., Corradi, Z., Khan, M., Hollander, A.I.D., Farrar, G.J., Gilissen, C., Hoischen, A., van de Vorst, M., et al. (2022). "Using single molecule Molecular Inversion Probes as a cost-effective, high-throughput sequencing approach to target all genes and loci associated with macular diseases." Hum. Mutat. *43*, 2234–2250. https://doi.org/10.1002/humu.24489.
45. Wolock, C.J., Stong, N., Ma, C.J., Nagasaki, T., Lee, W., Tsang, S.H., Kamalakaran, S., Goldstein, D.B., and Allikmets, R. (2019). A case-control collapsing analysis identifies retinal dystrophy genes associated with ophthalmic disease in patients with no pathogenic ABCA4 variants. Genet. Med. *21*, 2336–2344.
46. Runhart, E.H., Sangermano, R., Cornelis, S.S., Verheij, J.B.G.M., Plomp, A.S., Boon, C.J.F., Lugtenberg, D., Roosing, S., Bax, N.M., Blokland, E.A.W., et al. (2018). The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. Invest. Ophthalmol. Vis. Sci. *59*, 3220–3231.