Phenotype-Based Genetic Analysis Reveals Missing Heritability of *ABCA4*-Related Retinopathy: Deep Intronic Variants and Copy Number Variations

Lu Tian, Chunjie Chen, Yuning Song, Xiaohui Zhang, Ke Xu, Yue Xie, Zi-Bing Jin, and Yang Li

Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology & Visual Sciences Key Lab, Beijing, China

Correspondence: Yang Li, Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Hougou Lane 17, Chongwenmen Nei Street, Beijing 100730, China; yanglibio@aliyun.com.

Received: December 28, 2021
Accepted: May 19, 2022
Published: June 3, 2022

Citation: Tian L, Chen C, Song Y, et al. Phenotype-based genetic analysis reveals missing heritability of *ABCA4*-related retinopathy: Deep intronic variants and copy number variations. *Invest Ophthalmol Vis Sci.* 2022;63(6):5. https://doi.org/10.1167/iovs.63.6.5

Purpose. To identify the missing heritability of *ABCA4*-related retinopathy in a Chinese cohort.

Methods. We recruited 33 unrelated patients with *ABCA4*-related retinopathy carrying a monoallelic variant in *ABCA4*. All patients underwent ophthalmic examinations. Next-generation sequencing of the whole *ABCA4* sequence, including coding and noncoding regions, was performed to detect deep intronic variants (DIVs) and copy number variations (CNVs).

Results. We identified eight missing pathogenic *ABCA4* variants in 60.6% of the patients (20/33), which comprised five DIVs and three CNVs. The five DIVs, including four novel (c.1555-816T>G, c.2919-169T>G, c.2919-884G>T, and c.5461-1321A>G) and one reported (c.4539+1100A>G), accounted for the missing alleles in 51.5% of the patients. Minigene assays showed that four novel DIVs activated cryptic splice sites leading to the insertions of pseudoexons. The three novel CNVs consisted of one gross deletion of 1273 bp (exon 2) and two gross duplications covering 25.2 kb (exons 28–43) and 9.4 kb (exons 38–44). The microhomology domains were identified at the breakpoints and revealed the potential mechanisms of CNV formation.

Conclusions. DIVs and CNVs explained approximately two-thirds of the unresolved Chinese cases with *ABCA4*-related retinopathy. Combining results from phenotypic-directed screening, targeting the whole *ABCA4* sequencing and in silico tools can help to identify the missing heritability.

Keywords: *ABCA4*, deep intronic variant, copy number variation

The *ABCA4* gene (MIM 601691) is one of the most common disease-causing genes associated with inherited retinal degeneration and has a high carrier frequency of approximately 1:20.[1] *ABCA4* is located on chromosome 1p22.1 and contains 50 exons. It encodes the ABCA4 protein, which is located at the rim of the disk membranes in the outer segments of both the cone and rod photoreceptors of the human retina. The biallelic *ABCA4* variants cause a group of *ABCA4*-related retinopathies, including Stargardt disease (STGD1),[2] cone–rod dystrophy (CRD),[3] retinitis pigmentosa,[4] generalized choriocapillaris dystrophy,[5] and rapid-onset chorioretinopathy.[6]

The extensive clinical heterogeneity of *ABCA4*-related retinopathy arises mainly owing to the high genetic heterogeneity of the *ABCA4* gene.[1] Currently, more than 1600 disease-causing *ABCA4* variants have been identified, but the correlation between genotype and phenotype has not been explained fully. In previous studies, a genotype–phenotype model was proposed based on the putative functional consequences of the combination of *ABCA4* alleles.[7] The *ABCA4* variants were classified as "mild," "moderate," and "severe," and the random combinations of two variant alleles led to different phenotypes.[1,7,8] A combination of two severe *ABCA4* alleles usually results in severe forms of CRD.[1]

Sequencing of the coding regions and flanking splice sites has identified biallelic *ABCA4* variants in about 60% to 70% of *ABCA4*-related retinopathy cases,[9–11] and the remaining cases include patients with a single heterozygous variant (20%–25%) or no variant detected (10%–15%).[10,12] In 2013, Braun et al.[13] published the first report of pathogenic deep intronic variants (DIVs) in *ABCA4*. Since then, more than 30 DIVs have been found in *ABCA4*,[10,14–19] and may account for as much as 5% of all *ABCA4* pathogenic alleles.[1] Consequently, *ABCA4* has become the gene with the greatest number of described DIVs.

Current estimates indicate that approximately 10% of *ABCA4*-related retinopathy probands carry a DIV.[1] The results from in vitro splice assays or patient-derived cells suggest that the DIVs often lead to the insertions of an aberrant pseudoexon (PE) in several ways, such as by strengthening or creating an exonic splicing enhancer and intronic splicing silencer, by weakening or destroying an

exonic splicing silencer and intronic splicing enhancer, or by acting on cis-regulatory elements.[20] These insertions can be prevented by antisense oligonucleotide rescue. The promising effects of AONs bring new hope to patients with *ABCA4*-related retinopathy who carry DIVs in *ABCA4*.[17,21] Previous studies have identified copy number variations (CNVs) only in a minor fraction of the patients.[22–25]

In the current study, the whole *ABCA4* sequence was screened thoroughly using next-generation sequencing based on the clinical phenotypes of a cohort of Chinese patients carrying a monoallelic variant. The overall goal was to identify the DIVs and CNVs in *ABCA4*.

Methods

Recruitment of Subjects

A total of 33 unrelated Chinese patients were enrolled from 278 probands diagnosed with *ABCA4*-related retinopathy, who were from the Genetics Laboratory of the Beijing Institute of Ophthalmology, Beijing Tongren Ophthalmic Center, from 2008 to 2021. The inclusion criteria for patients were as follows: (1) clinical phenotypes were consistent with *ABCA4*-related retinopathy, and (2) only a monoallelic variant in *ABCA4* (NM_00350.2) was identified in the coding regions and flanking splice sites. This study strictly followed the principles of the Declaration of Helsinki and was approved by the Joint Committee of Clinical Research and the Ethics Committee of Beijing Tongren Hospital.

Clinical Evaluations

After obtaining their informed consent, all participants underwent detailed ophthalmic examinations, including best-corrected visual acuity with E decimal charts, slit-lamp biomicroscopy, and fundus examination and photography. Most patients also underwent optical coherence tomography, full-field electroretinography, and fundus autofluorescence. The diagnosis of STGD1 and CRD and the qualitative electroretinography assessment were based on previously reported criteria.[11,26] All probands were evaluated by qualified retina specialists.

Next-Generation Sequencing of Whole *ABCA4* Sequence and Data Analysis

Extracting genomic DNA and sequencing the coding regions and exon–intron boundaries of *ABCA4* was completed in preliminary work and some of the data were published.[11] Each proband showed one pathogenic allele in the coding regions or the exon–intron boundaries. Here, next-generation sequencing was applied to detect the undiscoverable variants in *ABCA4*. Targeted deep sequencing with the Agilent SureSelectXT kit (Agilent, Santa Clara, CA) was used to enrich the whole *ABCA4* sequence (chr1:94458240-94586850, GrCh37/hg19 assembly), with a minimum of 95.7% coverage to at least 500×. Samples were sequenced with a NovaSeq 6000 sequencing platform in the paired-end mode, according to the manufacturer's protocol. Sequencing data were analyzed using ANNOVAR software (version 2018-04-16, Qiagen Bioinformatics, Germany), and the reads were compared with the reference sequence of GRCh37 from the UCSC Genome Browser. A comparison was also conducted using the genome aggregation database (gnomAD, https://gnomad.broadinstitute.org/, last accessed March 14, 2021) to exclude nonpathogenic polymorphisms. A variant was classified as benign if the minor allele frequency was 0.05 or greater.

Five algorithms, including Human Splicing Finder (version 2.4.1), Splice Site Finder-like (from the Alamut Visual Plus 1.0), MaxEntScan (version 2003-7-22), NetGene2 (version 2.42), and NNSplice (version 0.9), were used to assess the possibility that noncoding region variants would induce aberrant transcript splicing. Prediction scores were provided for the wild type (WT) and the mutant type (MUT). Variants were selected for further testing by in vitro minigene assays when they were located at putative splice sites with a relative strength of at least 60% of the maximal score in at least two out of five splice prediction programs, or when they led to an increased splice prediction scores of at least 30% compared with the WT.

CNVs were analyzed preliminarily from variations in the read depth using CNV kit software. Assumed deletions or assumed duplications were considered candidate regions when values were less than 0.6 or greater than 1.4, respectively. Sequencing reads using Integrative Genomics Viewer software were examined to locate the breakpoints and Sanger sequencing was performed across the predicted breakpoints to confirm the breakpoint junctions. Tks Gflex DNA Polymerase (TaKaRa, Tokyo, Japan) was used for long-range PCR. The sequences of all primers are listed in Table 1. The potential mechanisms of the CNVs were assessed using in silico tools. Homology at the breakpoints was examined by multiple sequence comparison by log-expectation. The Dfam database was used to search for the transposable elements near the breakpoints. If both breakpoints of a CNV overlapped with a transposable element, the sequence identity between the transposable elements was determined using BLAST 2. The potential for the formation of non-B DNA conformations in the breakpoint regions was analyzed using the non-B DNA motif search tool.

Variants were classified as recommended by the American College of Medical Genetics and Genomics standards.

Table 1. *ABCA4* Primers to Amplify Three CNV Fragments for Breakpoints

CNV	Forward Primer Sequence	Reverse Primer Sequence	Region	T (°C)
c.67-140_161-355del	5′-AACAAGGTAGAGGGTGTATGGTGG-3′	5′-GCTCAGCAAAGCCACAAGAACACT-3′	Proximal, junction, distal	60
c.4128+246_6006-716dup	5′-GGCATTAGAGATCCAGACCTTATAGGCA-3′	5′-TCTCTCATTGGTGAAGGTCCCAGT-3′	Proximal	58
	5′-TTCTTGATCTCTAGGGCCAGGCTA-3′	5′-TCTCTCATTGGTGAAGGTCCCAGT-3′	Junction	62
	5′-TTCTTGATCTCTAGGGCCAGGCTA-3′	5′-GCACGCTTCAGTTTCTCATCTCCA-3′	Distal	62
c.5313-85_6147+21dup	5′-AGTCGAATCAGAATCCCAGACCCT-3′	5′-ACTCATCCGCATACAGCTGCTACA-3′	proximal	62
	5′-AAGCTTCTCCAGCCCTAGCTCTAT-3′	5′-ACTCATCCGCATACAGCTGCTACA-3′	Junction	60
	5′-AAGCTTCTCCAGCCCTAGCTCTAT-3′	5′-TGCAGTTTCTAAATCTGCCGACCC-3′	Distal	58

Abbreviations: del, deletion; dup, duplication; T, melting temperature.

Investigative Ophthalmology & Visual Science

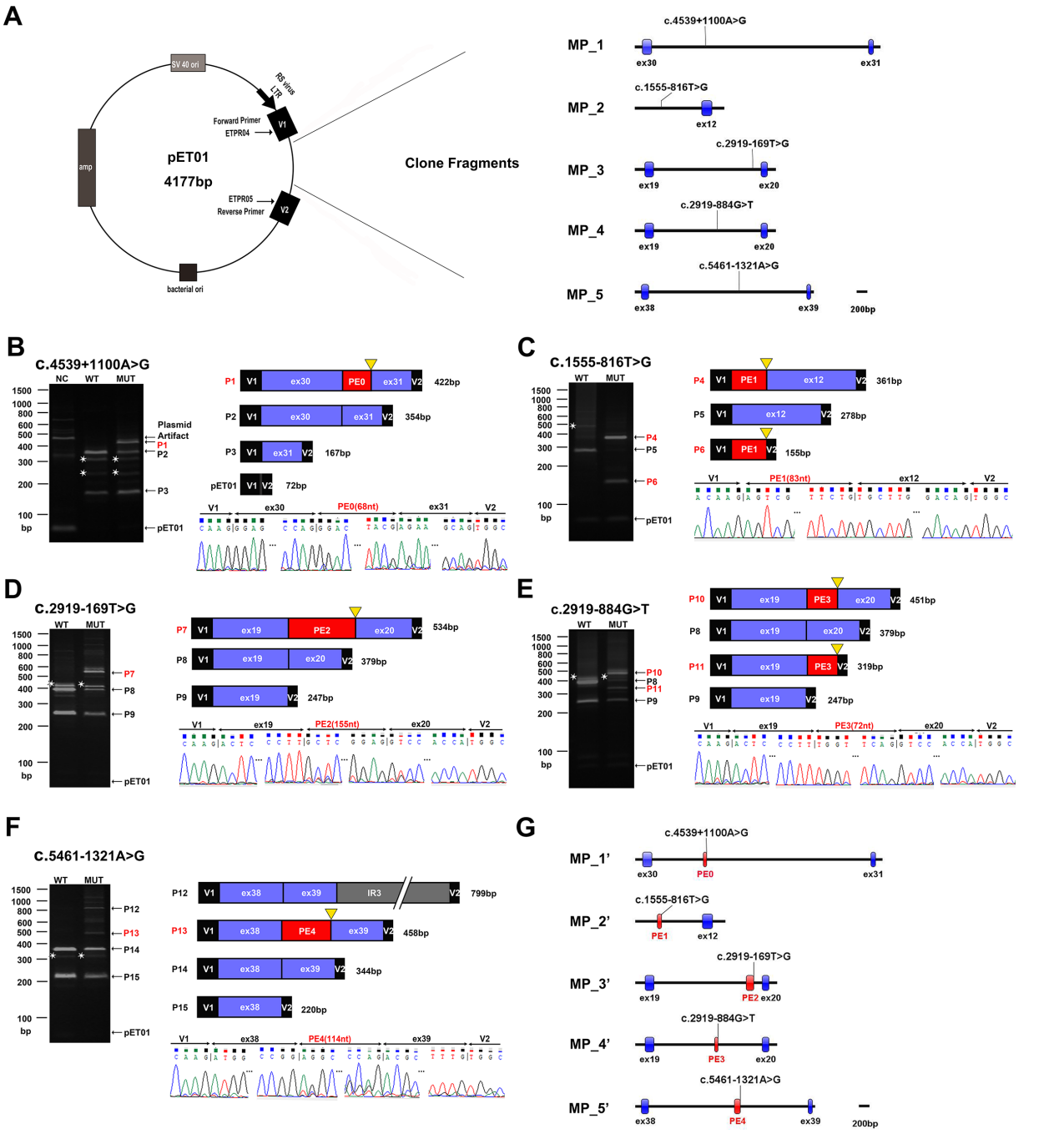

Figure 1. Overview of splice defects owing to four DIVs and one canonical splice site variant in *ABCA4*. All WT and MUT minigenes were transfected into ARPE-19 cells and their RNA was subjected to RT-PCR. The cDNA was amplified by using primers ETPR04 and ETPR05 in the exons of pET01. (**A**) Schematic representation of the five groups of minigenes used that were cloned between exons V1 and V2 in pET01. Except for the group MP_1, which only included the MUT minigene, the other four groups included WT minigenes and MUT minigenes. Positions of the variants present in the five minigenes are indicated. (**B**) RT-PCR for the negative control (empty pET01) and the positive control (c.4539+1100A>G MP_1 minigene) showed splicing defects. The negative control showed the normal splicing of exons V1 and V2. Three defects (fragments P1, P2, and P3) were observed in the positive control. One defect (fragment P1) was observed next to the WT fragments. A 68-nt PE inclusion was detected in the MUT constructs when compared with the WT. (**C**) RT-PCR products of MP_2 containing c.1555-816T>G variant residing in the deep intron 11. Two defects (fragments P4 and P6) were observed next to the WT fragments. An 83-nt PE inclusion was detected in the MUT constructs when compared with the WT. (**D**) RT-PCR products of MP_3 containing c.2919-169T>G variant residing in the deep intron 19. One defect (fragments P7) was observed next to the WT fragments. A 155-nt PE inclusion was detected in the MUT constructs compared with the WT. (**E**) RT-PCR products of MP_4 containing c.2919-884G>T variant residing in the deep intron 19. Two defects (fragments P10 and P11) were observed next to the WT fragments. A 72-nt PE inclusion was detected in the MUT constructs compared with the WT. (**F**) RT-PCR products of MP_5 containing the c.5461-1321A>G variant residing in the deep intron 38. One defect (fragments P13) was observed next to the WT fragments. A 114-nt PE inclusion was detected in the MUT constructs when compared with the WT. (**G**) Schematic representation of the five group minigenes, showing the positions of the variants and the PEs. Each of the four variants produced a new donor splice site. IR, intron retention; NC, negative control; P, product; PC, positive control. Asterisk denotes the fragment for which no sequence information was obtained. The triangle points to the position of the variant.

Investigative Ophthalmology & Visual Science

Table 2. *ABCA4* Primers to Amplify Genomic Fragments of Five Groups for Minigene Assays

Fragment	Exon	Size (bp)	Forward Primer Sequence	Reverse Primer Sequence	Restriction Enzyme	T (°C)
MP_1	30, 31	4959	5′-GCGCACTAGTCAGAAACCAGCCTCCAAGAG-3′	5′-GCGCGGCCGCGGGCCCTCAAATCAGATAAA-3′	Spe1-Not1	60
MP_2	12	2246	5′-GCGCACTAGTACAAGATATGGATGCCCTGG-3′	5′-GCGCGGCCGCAAGCAGGTTCCTCAGTCAGA-3′	Spe1-Not1	66
MP_3	19, 20	2843	5′-GCGCACTAGTTGGCATATAGGAGGTGCTGTATAA-3′	5′-GCGCGGCCGCCACTGCAAGTAGGACAAAGGACT-3′	Spe1-Not1	60
MP_4	19, 20	2843	5′-GCGCACTAGTTGGCATATAGGAGGTGCTGTATAA-3′	5′-GCGCGGCCGCCACTGCAAGTAGGACAAAGGACT-3′	Spe1-Not1	60
MP_5	38, 39	3607	5′-GCGCGGATCCGGGACTCTGGCCAGTTCACACAC-3′	5′-GCGCGGCCGCGGCACCCTAATCCTCTCCAGCT-3′	BamH1-Not1	60

The underlined text indicates the restriction site.

Minigene Assay in Human Retinal Pigmented Epithelium-19 (ARPE-19) Cells

In this study, four novel DIVs were highly suspected to cause splicing abnormalities. The four DIVs and one reported DIV c.4539+1100A>G (as a positive control) were evaluated for an effect on splicing with the pET01-based exon trapping system (Exontrap, MoBiTec GmbH, Goettingen, Germany). Except for the c.1555-816T>G variant, sequences including each DIV and exons (upstream and downstream) were PCR amplified from genomic DNA using different oligonucleotide primer pairs (minigene plasmid [MP]_1, MP_3, MP_4, and MP_5 in Fig. 1A). For the c.1555-816T>G variant located in intron 11, the cloned fragment only included exon 12 and part of the intron 11 sequence (MP_2 in Fig. 1A), because intron 11 is too long to be cloned into the pET01 vector. The 5′ end of the forward primers was designed to include a BamHI site or SpeI site, and the 5′ end of the reverse primers included a NotI site (Table 2). The amplified products were cloned into the pET01 vector. Clones with WT or MUT alleles were detected and screened by Sanger sequencing.

The authentication of ARPE-19 cells was verified by short tandem repeat analysis (Supplementary Fig. S1). The cells were transfected with 3 µg of the selected MP using a lipofectamine 2000 DNA transfection reagent (Invitrogen, Carlsbad, CA). After 48 h, the transfected cells were harvested, and total RNA was extracted using an EndoFree Plasmid Maxi Kit (Qiagen, Germany). The RT-PCR was performed with a pair of specific primers (ETPR04 and ETPR05 in the exons V1 and V2 of pET01) and the FastKing One-Step RT-PCR Kit (Tiangen BiotechBeijing, , China). The products were separated by electrophoresis on 2% agarose gels, excised, and sequenced.

Results

Molecular Diagnosis

The current cohort included 28 patients diagnosed with STGD, and five patients diagnosed with CRD or COD (cone dystrophy) (Supplementary Table S1). The typical phenotypes of the STGD and CRD are shown in Supplementary Figure S2. We identified eight missing pathogenic *ABCA4* variants in 20 patients (60.6%, 20/33); the missing alleles of the other 13 patients remained unresolved (Supplementary Table S1, Supplementary Fig. S3). The identified missing pathogenic *ABCA4* variants included five DIVs and three novel CNVs (Table 3).

Except for c.4539+1100A>G,[17] four DIVs (c.1555-816T>G, c.2919-169T>G, c.2919-884G>T, and c.5461-1321A>G) were novel. Totally, the five DIVs account for the missing alleles in 51.5% (17/33) of patients who carried a monoallelic variant of *ABCA4* and represented 25.8% (17/66) of all screened alleles. The variant with the highest allele frequency was c.4539+1100A>G identified in 7 (7/66, 10.6%) probands, followed by the novel DIV c.5461-1321A>G with the second-highest allele frequency (6/66, 9.1%), while c.2919-884G>T ranked the third with an allele frequency of 3.0% (2/66). The remaining two novel DIVs, c.1555-816T>G and c.2919-169T>G each, were each identified in one patient with an allele frequency of 1.5% (1/66).

The three CNVs identified in this cohort were novel. They consisted of one gross deletion and two gross duplications and accounted for the missing allele in 9.1% (3/33), representing 4.5% (3/66) of all screened alleles. The gross deletion c.67-140_161-355del was identified in patient 010626 as heterozygous (Fig. 2). The deletion, encompassing 1273 bp (exon 2), was inherited from her father. One gross duplication c.4128+246_6006-716dup, found in patient 010278, ranged from chr1:94471854 to 94497088 and covered 25,235 bp (exons 28–43) (Fig. 3). And another duplication c.5313-85_6147+21dup, found in patient 010658, involved 9356 bp (exons 38–44) (Fig. 4).

In total, this work revealed 41 variants in the 33 probands who carried an *ABCA4* monoallelic variant, including the aforementioned 8 missing pathogenic variants and 33 variants located in the coding regions or exon–intron boundaries. The 33 variants included 24 missense, three nonsense, three small deletions or insertions, and three splicing variants. Of the 33 variants,

Table 3. Overview of the DIVs and CNVs Identified in This Study

Genomic Position (hg19, chr1)	DNA Variant	RNA Effect	Protein Effect	Size	Patient
94529689	c.1555-816T>G	r.(1554_1555ins1555-898_1555-816)	p.(Cys519Serfs*77)	83 nt	010672
94510469	c.2919-169T>G	r.[=,2918_2919ins2919-323_2919-169]	p.[=,Ser974Leufs*5]	155 nt	010174
94511184	c.2919-884G>T	r.[2918_2919ins2919-957_2919-886,=]	p.[Phe973Leufs*3,=]	72 nt	010444 010708
94493901	c.4539+1100A>G	r.[4539_4540ins4539+1033_4539+1100,=]	p.[Arg1514Glyfs*3,=]	68 nt	010091 010105 010493 010506 010565 010611 010705
94478262	c.5461-1321A>G	r.[=,5460_5461ins5461-1435_5461-1322]	p.[=,Thr1821Argfs*18]	114 nt	010088 010093 010208 010554 010656 010716
94577490-94578762	c.67-140_161-355del	r.(67_161)del	p.(Ile23Alafs*24)	1273 bp	010626
94471854-94497088	c.4128+246_ 6006-716dup	r.(4128_6006)dup	p.(Ser2002Argfs*13)	25235 bp	010278
94470976-94480331	c.5313-85_ 6147+21dup	r.(5313_6147)dup	p.(Val2050Metfs*15)	9356 bp	010658

Abbreviations: bp, base pair; del, deletion; dup, duplication; ins, insertion; nt, nucleotide.

Investigative Ophthalmology & Visual Science

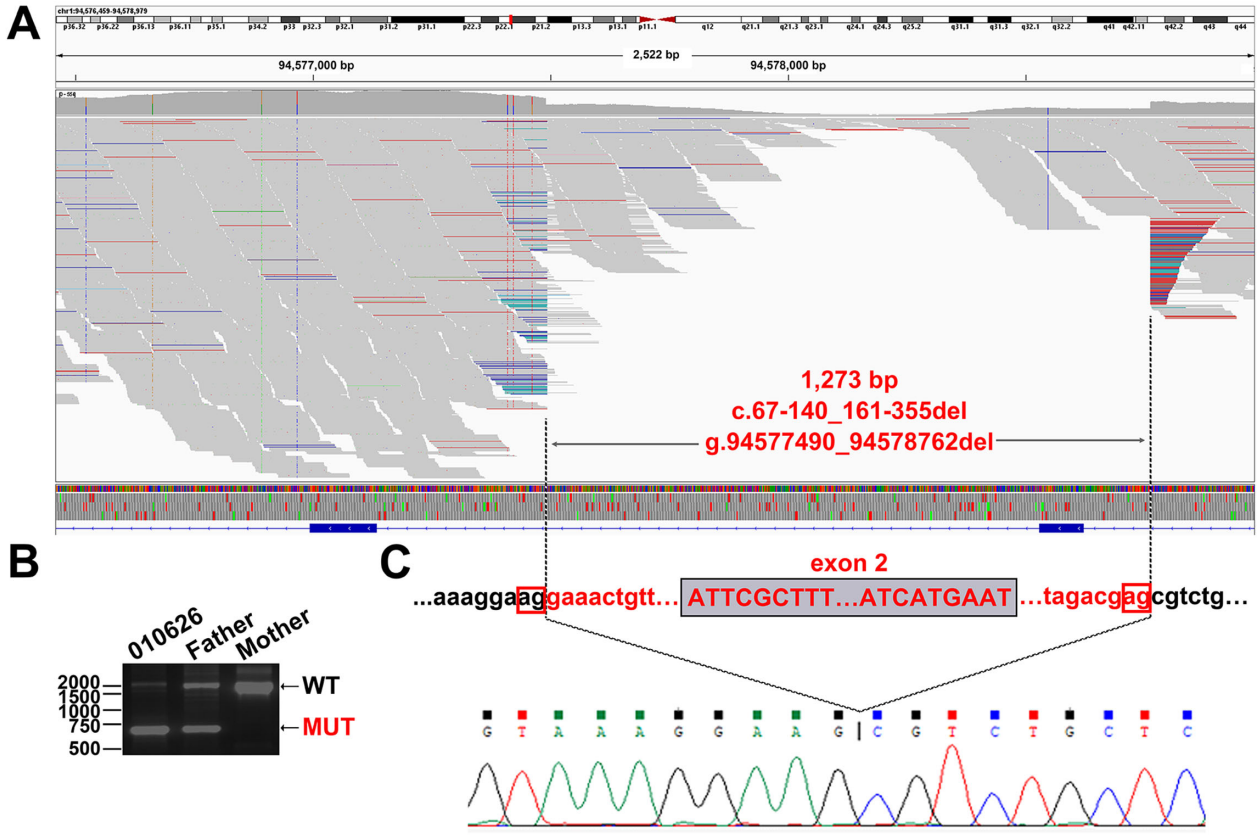

Figure 2. A gross deletion c.67-140_161-355del was identified in proband 010626 by next-generation sequencing of the whole *ABCA4* sequence. (**A**) The Integrative Genomics Viewer plot showed the deleted region and the breakpoints identified owing to the drop in coverage and increased distance between mate-paired reads. (**B**) The large deletion, which was inherited from her father, was confirmed by PCR. (**C**) Sanger sequencing also confirmed the deletion region from chr1:g.94577490 to g.94578762. The red rectangle indicates the repeated bases.

4 were not reported previously (Supplementary Table S2).[2,9,11,19,27–34]

Four Novel DIVs Validated by Minigene Assay

We summarized the scores of the splice predictors for each DIV and four novel DIVs met the inclusion criteria of the minigene assay (Supplementary Table S3). Overall, RT-PCR products showed that all four DIVs resulted in abnormal splicing (Fig. 1 and Supplementary Fig. S4). Bioinformatic analyses indicated that all DIVs activated the cryptic splice donor sites and led to the insertions of PEs. The abnormal splicing products were predicted to cause a premature termination codon, which leads to an unstable or nonfunctional protein or the degradation of the transcript by the nonsense-mediated decay process.

In the positive control group containing c.4539+1100A>G, the RT-PCR products showed three bands (P1, P2, and P3) in different sizes. The abnormal product P1 (422 bp) included the insertion of PE0 (68-nt) owing to the activation of the cryptic donor site p.[Arg1514Glyfs*3,=]. Multiple splicing products were observed both in WT construct and MUT constructs, including the P2 (354 bp) retaining the upstream and downstream exons, P3 (167 bp) retaining the downstream exon, and several bands without sequence. In addition, the empty plasmid produced an abnormal product which showed the insertion of the 5' upstream sequence of exon V1. This might be an experimental artifact caused by multiple replications of the plasmid.

In the MP_2 group, the mutant c.1555-816T>G construct showed two abnormal bands, P4 (361 bp) and P6 (155 bp), which lacked the correct splicing product (P5) (Fig. 1.C). Both P4 and P6 contained an 83-nt PE (PE1), which was predicted to form a stop codon at the 77th downstream amino acid p.(Cys519Serfs*77). The c.1555-816T>G significantly altered the motif's ratio of exonic splicing enhancers and exonic splicing silencers, thereby activating a cryptic intronic splice donor site (Fig. 1G). The activated cryptic splice donor site combined with a strong splice acceptor site situated 82-nt upstream resulted in the insertion of an 83-nt PE.

In the MP_3 group, the mutant c.2919-169T>G construct showed an additional band P7 (534 bp) when compared with the WT construct (P8, P9) (Fig. 1D). P7 contained a 155-nt PE (PE2) predicted to form a stop codon at the fifth downstream amino acid p.[=,Ser974Leufs*5]. The c.2919-169T>G was predicted to activate cryptic donor sites and was accompanied by strong splice acceptor sites situated 154-nt upstream (Fig. 1G). In addition to the correct products P8 and P9, a high-density band without an accurate sequence owing to interference from others was also found

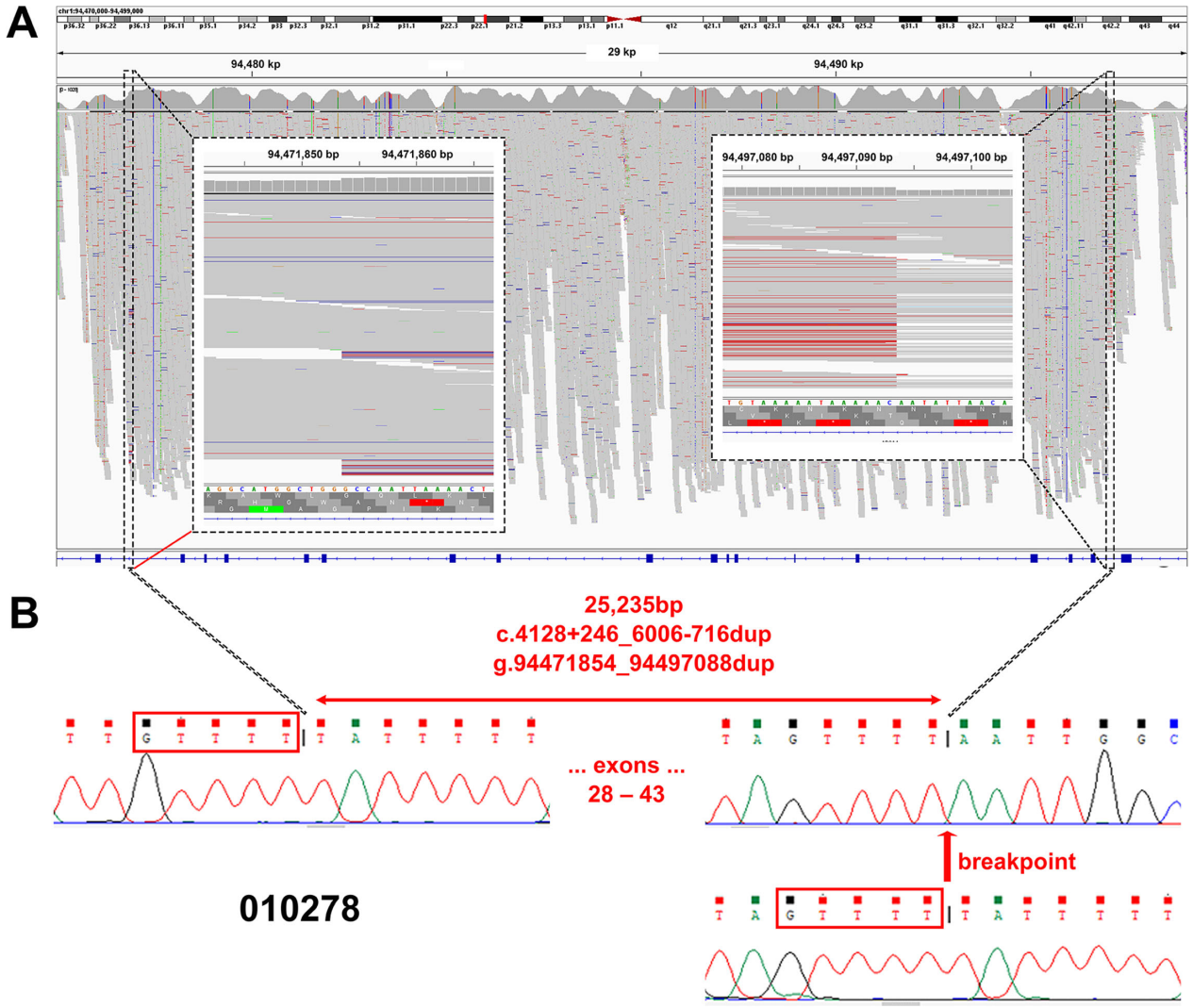

Figure 3. The gross duplication c.4128+246_6006-716dup was identified in proband 010278 by next-generation sequencing of the whole *ABCA4* sequence. (**A**) The Integrative Genomics Viewer plot showed the duplication region and the breakpoint, which were identified owing to the sharp increase in coverage and the increase in truncated reads. (**B**) The duplication was confirmed by PCR. The red arrow points to the breakpoint. The red rectangle indicates the repeated bases.

both in the WT and mutant constructs (marked with an asterisk in Fig. 1D), suggesting that this band belonged to the correct products.

In the MP_4 group, the mutant c.2919-884G>T construct showed two additional bands P10 (451 bp) and P11 (319 bp) that differed from the products P8 and P9 (Fig. 1E). Both P10 and P11 contained a 72-nt PE (PE3) predicted to cause a stop codon at the third downstream amino acid p.[Phe973Leufs*3,=]. The c.2919-884G>T activated the cryptic donor site that combined with a strong splice acceptor site situated 71-nt upstream (Fig. 1G). As with the MP_3 group, the band marked with an asterisk was found in both the WT and mutant constructs.

In the MP_5 group, the mutant c.5461-1321A>G construct showed an additional band P13 (458 bp) when compared with the WT construct (P14 and P15) (Fig. 1F). P13 contained a 114-nt PE (P4) predicted to form a stop codon at the 18th downstream amino acid p.[Thr1821Argfs*18,=] (Fig. 1G). The c.5461-1321A>G variant activated a cryptic donor site by creating a new exonic splicing enhancers (SF2/ASF) motif and combined with a strong splicing acceptor site situated 113-nt upstream. The band P12 (799 bp) from the MUT construct retained the downstream intron, regarded as an experimental artifact arising from the downstream exon and part of the intron provided by the plasmid.

CNV Analysis

The breakpoints of the three novel CNVs were all validated by Sanger sequencing (Figs. 2–4). Bioinformatics analyses revealed microhomology at the breakpoints and the potential mechanisms of the three CNVs (Table 4, Supplementary Table S4). For the deletion c.67-140_161-355del, a 2-bp microhomology domain "AG" was identified at the breakpoint. Repeated elements MER5A1 and L2b_3end were found at both ends of the breakpoint, but no similar sequence was found between them after sequence identity analysis. For the gross duplication c.4128+246_6006-716dup, a 5-bp microhomology domain "GTTTT" was

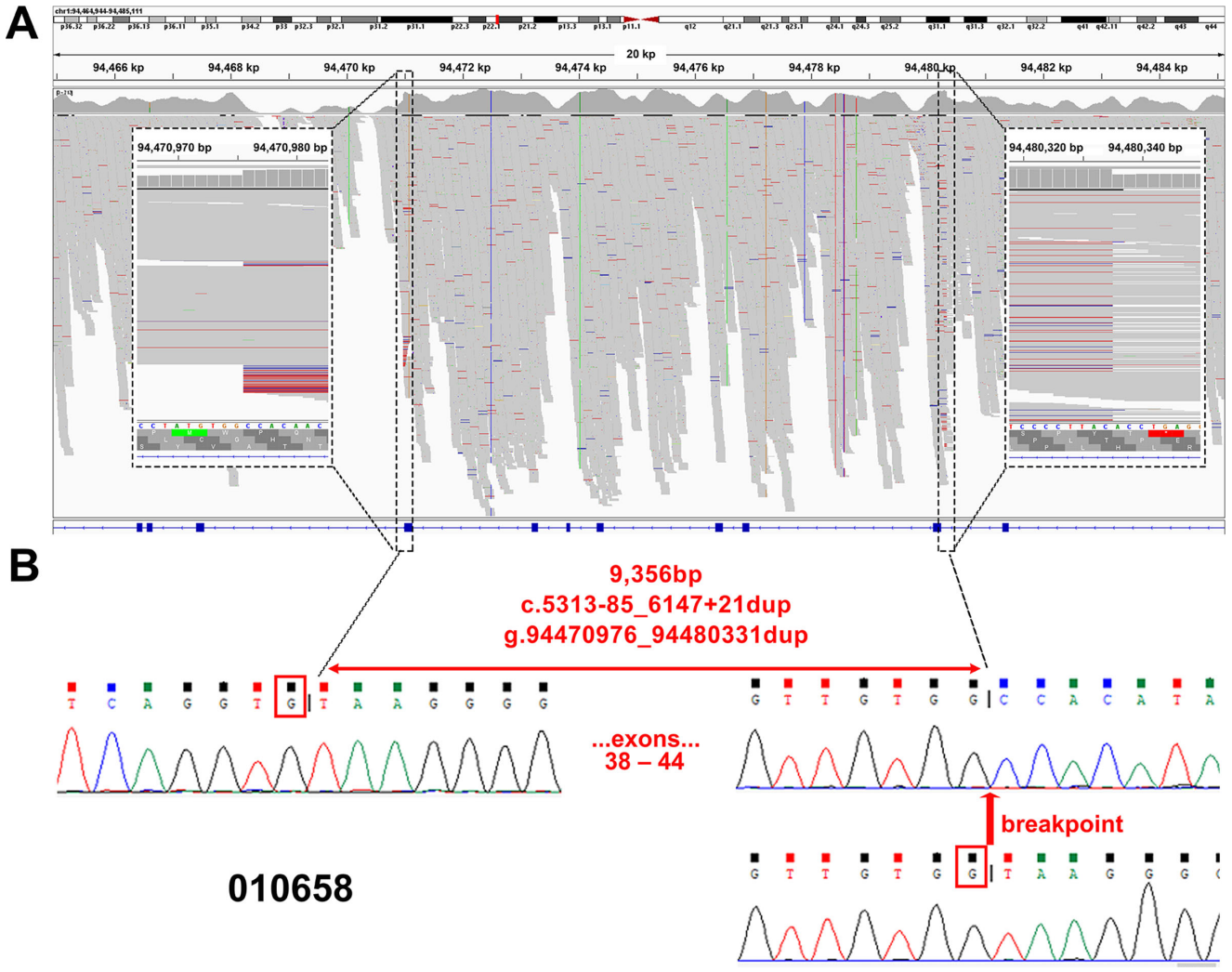

Figure 4. The gross duplication c.5313-85_6147+21dup was identified in proband 010658 by next-generation sequencing of the whole *ABCA4* sequence. (**A**) The Integrative Genomics Viewer plot showed the duplication region and the breakpoint which were identified owing to the sharp rise in coverage and the increase in truncated reads. (**B**) The duplication was confirmed by PCR. The red arrow points to the breakpoint. The red rectangle indicates the repeated base.

confirmed by Sanger sequencing. No repeated sequence or non-B DNA forming motif was found. For the gross duplication c.5313-85_6147+21dup, a 1-bp microhomology domain "G" was confirmed by Sanger sequencing. Repeated elements MIRc and L2a_3end were found at both ends of the breakpoint, without similar sequences between the two repeated elements. One mirror repeat (triplex motif) was detected at the 5′ end and spanned the breakpoint.

Discussion

This comprehensive genetic analysis is the first to examine the missing heritability of *ABCA4*-related retinopathy in a cohort of Chinese patients. Overall, we have unraveled the missing heritability, including DIVs and CNVs, in 60.6% of the patients who carried the *ABCA4* monoallelic variant in the current cohort. This percentage is higher than the data previously found (40.3%) in Caucasian patients.[21] In our cohort, DIVs accounted for approximately one-half of the missing alleles, slightly lower than the proportion (66.7%) reported by Sangermano et al.[17] CNVs accounted for nearly 10% of the missing alleles, similar to the proportion reported by Bauwens et al.[21]

Consistent with our previous descriptions for the disease-causing variant spectrum located in the coding regions and exon–intron boundaries of *ABCA4*, our results showed that the DIV spectrum observed in Chinese patients was quite different from that described in other populations.[11] The reported DIV c.4539+1100A>G was detected in 21.2% of the cases (7/33) in our cohort, which was a much higher proportion than the 5.6% (2/36) identified in Caucasian patients.[17] The second common DIV c.5461-1321A>G was identified in 18.2% cases (6/33) in our cohort and has not been found previously in other populations. By contrast, a founder DIV c.4539+2001G>A detected in more than 25% of Belgian patients with a monoallelic STGD1 was not detected in our cohort patients.[35] In addition, 80% of the reported DIVs

Table 4. The Microhomology of the Breakpoints of the Three Novel CNVs Assessed by MUSCLE

CNV	Region	Sequence
c.67-140_161-355del	Proximal	5′-**CCAGATGGACACAAAATCTGGTTGATATAATAAAACAAGGTAGAGGGTGTATGGTGGGG AGGGGGTAAAGGAAG**GAAACTGTTTAGGTAAAGATACCACAACCAAAGTCCTACTGCACACA TGGGATCTGAGGAGGGCTGTGTCTGCTCT-3′
	Junction	5′-**CCAGATGGACACAAAATCTGGTTGATATAATAAAACAAGGTAGAGGGTGTATGGTGGGG AGGGGGTAAAGGAAGCGTCTGCTCAAGCAGCAGCAGCAACTGCGTGGAGTCTTCTTGAA CTAACACTCCTATGCCCCTCTCGGCACAAAAT**-3′
	Distal	5′-GAGCCCATAACCTGCAGCCCTACTGCATGCTTGGGGTAGGTGCTCAGTTCACCGTGGTTGAA GGAATAGACG**AGCGTCTGCTCAAGCAGCAGCAGCAACTGCGTGGAGTCTTCTTGAACTAA CACTCCTATGCCCCTCTCGGCACAAAAT**-3′
c.4128+246_ 6006-716dup	Proximal	5′-ATGAATCTTCTTTGTAAGTTGGACCTCCCCTAGCAAAGAAAATAGAATAATAGTGAAAATGTTA ATATT**GTTTTTATTTTTACAGTGAGGGATAAAGTCATGTTTTCATTCATTTTTGCAGTGACC CTACATATCAAAATCATTGCCCTC**-3′
	Junction	5′-**AGAATTGGGATTATATGGTAGGGTAGCTCCCACTAATTTGGAAACGTACCCTACTTGCTTC CCTGAGTAGTTTTTATTTTTACAGTGAGGGATAAAGTCATGTTTTCATTCATTTTTGCAGT GACCCTACATATCAAAATCATTGCCCTC**-3′
	Distal	5′-**AGAATTGGGATTATATGGTAGGGTAGCTCCCACTAATTTGGAAACGTACCCTACTTGCTTC CCTGAGTAGTTTT**AATTGGCCCAGCCATGCCTTTGGTGGCTTTTGTCATTGTGGGGAACTGT AATGGTCTCTCTGTACCATCCTATATC-3′
c.5313-85_ 6147+21dup	Proximal	5′-CACAGAGACCAACAAATGGAGTATCTCTCTGGTCCCCTGGGACTCTGGCCAGTTCACACACAT CACCTCAGGT**GTAAGGGGAGTGCATTATATCCAGACGTATTGTAGGTGGAATGGAATGTG GAACTCCATCACTCTGAGTTGTCTCAT**-3′
	Junction	5′-**ATCTTTACCTTTATGCCCGGCTTCGAGGTGTACCAGCAGAAGAAATCGAAAAGGTGAAA AATGTTTTGTTGTGGTAAGGGGAGTGCATTATATCCAGACGTATTGTAGGTGGAATGGAA TGTGGAACTCCATCACTCTGAGTTGTCTCAT**-3′
	Distal	5′-**ATCTTTACCTTTATGCCCGGCTTCGAGGTGTACCAGCAGAAGAAATCGAAAAGGTGAAA AATGTTTTGTTGTGG**CCACATAGGAGTCTGGTTAATTACAAGCCTGTTTCATGAGAGTGCAT TCTCTTGGAGATGAGAAACTGAAGCGTGC-3′

MUSCLE, multiple sequence comparison by log-expectation. The bold text indicates the sequence of the junction region, and the bold and underlined text indicates the microhomology domain.

(>30) were concentrated in the introns 6, 7, 13, 30, and 36, whereas the four novel DIVs were distributed in introns 11, 19, and 38.

Unlike the previous studies, we conducted an in vitro function analysis of the DIVs in ARPE-19 cells, rather than in the human embryonic kidney cells (HEK293T).[17,18,21,36] The ARPE-19 cells are derived from human RPE cells, so we considered that they might better recapitulate the retina-specific splicing factors.[37] Our results indicated that all the DIVs activated the cryptic splice donor sites to produce one or more abnormal splicing products containing an insertion of a PE. All the PE insertions altered the reading frame and led to a premature termination codon. Except for the DIV c.1555-816T>G, our minigene analysis revealed that the other four DIVs (including the reported c.4539+1100A>G) generated both normal and abnormal splicing products. Sangermano et al.[17] also observed both correct and mutant transcripts in the majority of the DIVs detected in *ABCA4*, but the fraction of mutant transcripts differed (from 8% to 93%) for each DIV. In the current study, DIV c.5461-1321A>G yielded the lowest density of mutant transcripts (the band P13), whereas DIV c.1555-816T>G produced the highest density of abnormal transcripts. However, we were unable to determine or predict their pathogenic strengths owing to the lack of quantitative analysis of the RT-PCR products. Moreover, the functional consequences of most of the DIVs in *ABCA4* have been verified using in vitro minigene assays; therefore, the findings may not fully reflect the in vivo situation in the retina. One recent study further elucidated the significance of retinal differentiation for PE recognition of the DIV c.2991+1655A>G in *CEP290*.[38] The lymphoblastoid and fibroblast cells of a patient carrying the homozygous DIV c.2991+1655A>G presented a 1:1 ratio between correctly and aberrantly spliced *CEP290*, but this ratio changed to 1:4 in photoreceptor cells derived from induced pluripotent stem cells. This variation in the splicing results from different cell types was also observed in two patients harboring the heterozygous DIVs, c.4539+2028C>T or c.4539+2001G>A in *ABCA4*.[17] The patient-derived fibroblasts only expressed the normal *ABCA4* transcript, while patient-derived photoreceptor precursor cells expressed both the normal and the mutant transcripts. Therefore, induced pluripotent stem cell–derived retinal cells or photoreceptor precursor cells from patients would be appropriate cellular systems for future studies on splice defects of DIVs.

DIVs activate or create new splice sites that lead to the insertions of PEs, in agreement with our results.[20] The identification of the pathogenic mechanism of DIVs has opened up a new treatment strategy for patients carrying these variants. Studies have shown that AONs manipulated the splicing process by blocking the insertions of PEs.[17,21,36] The drug QR-110, the first antisense oligonucleotide applied to treat hereditary retinopathy (*CEP290* associated with Leber's congenital amaurosis), has demonstrated encouraging safety and efficacy data in phase II clinical trials, and a phase III clinical trial is ongoing (https://clinicaltrials.gov/ct2/show/NCT03913143, last accessed June 6, 2021).[39]

In contrast with the DIVs, CNVs were very rare in the *ABCA4* gene,[22–25] representing less than 1% of all alleles. The combination of the next-generation sequencing data analysis and Sanger sequencing validated the breakpoints of the three novel CNVs reported here. Three microhomology domains were identified at the breakpoints and revealed the

potential mechanisms of CNV formation. The microhomology at the breakpoints suggested that the CNVs belonged to the nonrecurrent rearrangements and that the underlying mechanisms might involve the nonhomologous end joining or microhomology-mediated end joining.[40,41] Our results demonstrated that these breakpoints serve an essential role in resolving the potential mechanisms of CNV formation. Combining data from targeted sequencing and in silico tools was an effective and cost-effective method for analyzing CNV breakpoints.

In our cohort, 39.4% of the cases (13/33) remain unresolved, for several possible reasons. For example, the inclusion criteria for the splicing assay may have been too restrictive and may have excluded some pathogenic deep intronic *ABCA4* variants. The number of DIVs might increase when the screening criteria are relaxed, for example, by decreasing the inclusion score or increasing the algorithms. Alternatively, the remaining pathogenic *ABCA4* variants may be located in regulatory regions, such as promoter and enhancer sequences. Rare variants in these regions should be sought out by performing bioinformatic analyses and quantitative analyses of the ABCA4 protein. In addition, the inversions and insertions may have been difficult to detect by short-read sequencing strategies, and our coverage of the entire *ABCA4* sequence was not 100%. Long-read sequencing would be helpful for the comprehensive discovery of these variants.

Conclusions

Our results indicate that DIVs and CNVs can explain approximately two-thirds of unresolved cases with *ABCA4*-related retinopathy. Combining the results from phenotypic screening, targeted whole *ABCA4* sequencing and in silico tools can help to identify the missing heritability.

Acknowledgments

Supported by the National Key R&D Program of China, 2017YFA0104103. The funding organization had no role in the design or conduct of this research.

Disclosure: **L. Tian,** None; **C. Chen,** None; **Y. Song,** None; **X. Zhang,** None; **K. Xu,** None; **Y. Xie,** None; **Z.-B. Jin,** None; **Y. Li,** None

References

1. Cremers FPM, Lee W, Collin RWJ, Allikmets R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. *Prog Retin Eye Res*. 2020;79:100861.
2. Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*. 1997;15:236–246.
3. Maugeri A, Klevering BJ, Rohrschneider K, et al. Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet*. 2000;67:960–966.
4. Martinez-Mir A, Paloma E, Allikmets R, et al. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet*. 1998;18:11–12.
5. Bertelsen M, Zernant J, Larsen M, Duno M, Allikmets R, Rosenberg T. Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci*. 2014;55:2766–2776.
6. Tanaka K, Lee W, Zernant J, et al. The rapid-onset chorioretinopathy phenotype of ABCA4 disease. *Ophthalmology*. 2018;125:89–99.
7. Fujinami K, Zernant J, Chana RK, et al. Clinical and molecular characteristics of childhood-onset Stargardt disease. *Ophthalmology*. 2015;122:326–334.
8. Del Pozo-Valero M, Riveiro-Alvarez R, Blanco-Kelly F, et al. Genotype-phenotype correlations in a Spanish cohort of 506 families with biallelic ABCA4 pathogenic variants. *Am J Ophthalmol*. 2020;219:195–204.
9. Lewis RA, Shroyer NF, Singh N, et al. Genotype/phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet*. 1999;64:422–434.
10. Zernant J, Xie YA, Ayuso C, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet*. 2014;23:6797–6806.
11. Jiang F, Pan Z, Xu K, et al. Screening of ABCA4 gene in a Chinese cohort with Stargardt disease or cone-rod dystrophy with a report on 85 novel mutations. *Invest Ophthalmol Vis Sci*. 2016;57:145–152.
12. Zernant J, Lee W, Collison FT, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet*. 2017;54:404–412.
13. Braun TA, Mullins RF, Wagner AH, et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet*. 2013;22:5136–5145.
14. Sangermano R, Bax NM, Bauwens M, et al. Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T–>C mutation in Stargardt disease. *Ophthalmology*. 2016;123:1375–1385.
15. Nassisi M, Mohand-Said S, Andrieu C, et al. Prevalence of ABCA4 deep-intronic variants and related phenotype in an unsolved "one-hit" cohort with Stargardt disease. *Int J Mol Sci*. 2019;20:5053.
16. Khan M, Cornelis SS, Khan MI, et al. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. *Hum Mutat*. 2019;40:1749–1759.
17. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med*. 2019;21:1751–1760.
18. Schulz HL, Grassmann F, Kellner U, et al. 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*. 2017;58:394–403.
19. Paavo M, Zhao J, Kim HJ, et al. Mutations in GPR143/OA1 and ABCA4 inform interpretations of short-wavelength and near-infrared fundus autofluorescence. *Invest Ophthalmol Vis Sci*. 2018;59:2459–2469.
20. Weisschuh N, Buena-Atienza E, Wissinger B. Splicing mutations in inherited retinal diseases. *Prog Retin Eye Res*. 2021;80:100874.
21. Bauwens M, Garanto A, Sangermano R, et al. 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*. 2019;21:1761–1771.
22. Lambertus S, van Huet RA, Bax NM, et al. Early-onset Stargardt disease: phenotypic and genotypic characteristics. *Ophthalmology*. 2015;122:335–344.

23. Bax NM, Sangermano R, Roosing S, et al. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat*. 2015;36:43–47.
24. Carss KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet*. 2017;100:75–90.
25. Ellingford JM, Horn B, Campbell C, et al. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J Med Genet*. 2018;55:114–121.
26. Lois N, Holder GE, Bunce C, Fitzke FW, Bird AC. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol*. 2001;119:359–369.
27. Birtel J, Gliem M, Hess K, et al. Comprehensive geno- and phenotyping in a complex pedigree including four different inherited retinal dystrophies. *Genes (Basel)*. 2020;11:137.
28. Ducroq D, Rozet JM, Gerber S, et al. The ABCA4 gene in autosomal recessive cone-rod dystrophies. *Am J Hum Genet*. 2002;71:1480–1482.
29. Wiszniewski W, Zaremba CM, Yatsenko AN, et al. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Hum Mol Genet*. 2005;14:2769–2778.
30. Cella W, Greenstein VC, Zernant-Rajang J, et al. G1961E mutant allele in the Stargardt disease gene ABCA4 causes bull's eye maculopathy. *Exp Eye Res*. 2009;89:16–24.
31. Taylor RL, Parry NRA, Barton SJ, et al. Panel-based clinical genetic testing in 85 children with inherited retinal disease. *Ophthalmology*. 2017;124:985–991.
32. Testa F, Rossi S, Sodi A, et al. Correlation between photoreceptor layer integrity and visual function in patients with Stargardt disease: implications for gene therapy. *Invest Ophthalmol Vis Sci*. 2012;53:4409–4415.
33. Liu X, Fujinami K, Kuniyoshi K, et al. Clinical and genetic characteristics of 15 affected patients from 12 Japanese families with GUCY2D-associated retinal disorder. *Transl Vis Sci Technol*. 2020;9:2.
34. Hu FY, Li JK, Gao FJ, et al. ABCA4 gene screening in a Chinese Cohort with Stargardt disease: identification of 37 novel variants. *Front Genet*. 2019;10:773.
35. Bauwens M, De Zaeytijd J, Weisschuh N, et al. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat*. 2015;36:39–42.
36. Khan M, Arno G, Fakin A, et al. Detailed phenotyping and therapeutic strategies for intronic ABCA4 variants in Stargardt disease. *Mol Ther Nucleic Acids*. 2020;21:412–427.
37. Jonsson F, Westin IM, Osterman L, et al. ATP-binding cassette subfamily a, member 4 intronic variants c.4773+3A>G and c.5461-10T>C cause Stargardt disease due to defective splicing. *Acta Ophthalmol*. 2018;96:737–743.
38. Parfitt DA, Lane A, Ramsden CM, et al. Identification and correction of mechanisms underlying inherited blindness in human iPSC-derived optic cups. *Cell Stem Cell*. 2016;18:769–781.
39. Leroy BP, Birch DG, Duncan JL, et al. Leber congenital amaurosis due to CEP290 mutations severe vision impairment with a high unmet medical need: a review. *Retina*. 2021;41:898–907.
40. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*. 2016;17:224–238.
41. van den Akker J, Hon L, Ondov A, et al. Intronic breakpoint signatures enhance detection and characterization of clinically relevant germline structural variants. *J Mol Diagn*. 2021;23:612–629.

Investigative Ophthalmology & Visual Science