HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2021 June 07.

Published in final edited form as:

Hum Mutat. 2020 January ; 41(1): 255–264. doi:10.1002/humu.23920.

Deep-intronic variants in *CNGB3* cause achromatopsia by pseudoexon activation

Nicole Weisschuh[1], **Marc Sturm**[2], **Britta Baumann**[1], **Isabelle Audo**[3,4,5], **Carmen Ayuso**[6,7], **Beatrice Bocquet**[8,9], **Kari Branham**[10], **Brian P. Brooks**[11], **Jaume Catalá-Mora**[12], **Roberto Giorda**[13], **John R. Heckenlively**[10], **Robert B. Hufnagel**[11], **Samuel G. Jacobson**[14], **Ulrich Kellner**[15], **Sofia Kitsiou-Tzeli**[16], **Alexandre Matet**[17], **Loreto Martorell Sampol**[18], **Isabelle Meunier**[7,8], **Günther Rudolph**[19], **Dror Sharon**[20], **Katarina Stingl**[21], **Berthold Streubel**[22], **Balázs Varsányi**[23,24], **Bernd Wissinger**[1], **Susanne Kohl**[1]

[1]Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tübingen, Tübingen, Germany [2]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany [3]Institut de la Vision, Sorbonne Université, INSERM, CNRS, Paris, France [4]CHNO des Quinze-Vingts, INSERM-DHOS CIC1423, Paris, France [5]Institute of Ophthalmology, University College of London, London, United Kingdom [6]Department of Genetics, Instituto de Investigación Sanitaria–Fundación Jiménez Díaz University Hospital-Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain [7]Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain [8]Centre National de Référence «Maladies Sensorielles Génétiques», Service Ophtalmologie, Hôpital Gui de Chauliac, CHRU de Montpellier, Montpellier, France [9]INSERM U1051, Institute for Neurosciences of Montpellier, Montpellier, France [10]Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, Michigan [11]National Eye Institute, National Institutes of Health, Bethesda, Maryland [12]Ophthalmology, Hospital Sant Joan de Deu, Barcelona, Spain [13]Molecular Biology Lab, Scientific Institute, IRCCS Eugenio Medea, Bosisio Parini, Italy [14]Department of Ophthalmology, Perelman School of Medicine, Scheie Eye Institute, University of Pennsylvania, Philadelphia, Pennsylvania [15]Rare Retinal Disease Center, Augenzentrum Siegburg, MVZ ADTC Siegburg GmbH, Siegburg, Germany [16]Department of Medical Genetics, National & Kapodistrian University of Athens, Athens, Greece [17]Department of Ophthalmology, Jules-Gonin Eye Hospital, University of Lausanne, Lausanne, Switzerland [18]Laboratorio de Genética Molecular, Hospital Sant Joan de Deu, Barcelona, Spain [19]Department of Ophthalmology, Ludwig-Maximilians-University, Munich, Germany [20]Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel [21]University Eye Hospital, Center for Ophthalmology, University of Tübingen, Tübingen, Germany [22]Department of Pathology, Medical University of Vienna, Vienna, Austria [23]Department of Ophthalmology, Semmelweis University, Budapest, Hungary [24]Department of Ophthalmology, University of Pécs Medical School, Pécs, Hungary

Correspondence Nicole Weisschuh, Institute for Ophthalmic, Research, Centre for Ophthalmology, University of Tübingen, 72076 Tübingen, Germany., nicole.weisschuh@uni-tuebingen.de.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

CONFLICT OF INTEREST

The authors declare that there are no conflict of interest.

Abstract

Our comprehensive cohort of 1100 unrelated achromatopsia (ACHM) patients comprises a considerable number of cases (~5%) harboring only a single pathogenic variant in the major ACHM gene *CNGB3*. We sequenced the entire *CNGB3* locus in 33 of these patients to find a second variant which eventually explained the patients' phenotype. Forty-seven intronic *CNGB3* variants were identified in 28 subjects after a filtering step based on frequency and the exclusion of variants found *in cis* with pathogenic alleles. In a second step, in silico prediction tools were used to filter out those variants with little odds of being deleterious. This left three variants that were analyzed using heterologous splicing assays. Variant c.1663-1205G>A, found in 14 subjects, and variant c.1663-2137C>T, found in two subjects, were indeed shown to exert a splicing defect by causing pseudoexon insertion into the transcript. Subsequent screening of further unsolved *CNGB3* subjects identified four additional cases harboring the c.1663-1205G>A variant which makes it the eighth most frequent *CNGB3* variant in our cohort. Compound heterozygosity could be validated in ten cases. Our study demonstrates that whole gene sequencing can be a powerful approach to identify the second pathogenic allele in patients apparently harboring only one disease-causing variant.

Keywords

achromatopsia; CNGB3; deep intronic variant; pseudoexon; splicing defect

1 | INTRODUCTION

Achromatopsia (ACHM) is a rare autosomal recessive retinal disorder characterized by color vision defects, photophobia, nystagmus, and severely reduced visual acuity. The disease is caused by mutations in genes encoding crucial components of the cone phototransduction cascade, namely *CNGA3* (MIM# 600053), *CNGB3* (MIM# 605080), *GNAT2* (MIM# 139340), *PDE6C* (MIM# 600827), and *PDE6H* (MIM# 601190), or in ATF6 (MIM# 605537), involved in the unfolded protein response. *CNGB3* encodes the beta subunit of the cyclic nucleotide-gated ion channel in cone photoreceptors and is the major ACHM gene in Europe and the US. Yet a considerable number of cases remain genetically unsolved, in some cases because only a single pathogenic allele was identified for one of the known ACHM genes. We previously showed that copy number variations (CNVs) in *CNGB3* contribute to these missing alleles in *CNGB3*-linked ACHM (Mayer et al., 2017), but do not account for the second allele in all cases. From our entire ACHM cohort which comprises 1,100 independent families, we have selected 33 cases harboring a single disease-causing variant in *CNGB3* and in which CNV analysis failed to identify the second pathogenic allele. We hypothesized that the second pathogenic allele is a noncanonical splice site variant located in the intronic regions not covered by standard exon-oriented screenings. In fact, an important, but maybe underestimated cause of hereditary diseases is the inclusion of intronic sequences in mRNAs (Buratti et al., 2007; Vorechovsky, 2006). This pseudoexon inclusion is triggered by mutations that activate noncanonical splice sites (Vaz Drago et al., 2017). It was proposed that approximately 50% of pseudoexons are derived from transposable elements, particularly *Alu* elements (Lev-Maor, Sorek, Shomron, & Ast, 2003;

Vorechovsky, 2010). The genomic region of *CNGB3* is highly enriched with repeat elements as has been shown previously (Van Schil et al., 2018). The rationale behind our study was to identify intronic variants in the *CNGB3* gene that create novel acceptor or donor splice sites, thereby activating complementary splice sites and leading to the inclusion of pseudoexons in the mRNA. Whole-gene targeted sequencing of the *CNGB3* gene led to the identification of three distinct novel deep intronic variants that were predicted in silico to create cryptic donor splice sites. The consequences of these splicing alterations were validated using reporter minigene assays.

2 | MATERIALS AND METHODS

2.1 | Editorial policies and ethical considerations

Samples from all patients and family members were recruited in accordance with the principles of the Declaration of Helsinki and were obtained with written informed consent approved by the respective local research and ethical boards or dependent on the local regulatory bodies accompanying the patients' samples. Specifically, this study was approved by the institutional review board of the Ethics Committee of the University Hospital of Tübingen under the study no. 116/2015BO2.

2.2 | Subjects and clinical evaluation

The study cohort comprised 39 unrelated patients, recruited at several ophthalmic centers (from Germany, Switzerland, Spain, France, Denmark, Hungary, Greece, Austria, USA) and sent to the Institute for Ophthalmic Research in Tübingen (Germany) for genetic research investigation. The clinical diagnosis was established by standard clinical ophthalmologic examinations including, but not limited to, visual acuity, color vision, and ophthalmoscopy. Some cases underwent noninvasive retinal imaging, psychophysical and electrophysiological testing. Genomic DNA of patients was extracted from peripheral blood using standard protocols. Mutation screening for the six ACHM genes was performed as described previously (Mayer et al., 2017).

2.3 | Target enrichment

To cover the entire *CNGB3* gene (87,566,205–87,755,903 in GRCh37 coordinates), a set of 18 overlapping primer pairs (see Table S1) was designed using Primer3plus (Untergasser et al., 2007). The long distance (LD)-PCR approach covered 89.7% (155,849 bp) of 173,824 target basepairs. The uncovered bases represented 10.3% of the total number of bases, all of them located in noncoding regions and mostly comprising arrays of repeat elements such as Alu sequences, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Amplicons were generated using the LA Taq Polymerase (TaKaRa Bio, Saint-Germain-en-Laye, France) according to manufacturer's protocol. Amplification was verified by analyzing 2 μl of each PCR product on a 1% agarose gel. Quantification was performed by ImageJ (Schindelin, Rueden, Hiner, & Eliceiri, 2015) using five different concentrations of lambda DNA to generate the standard curve.

2.4 | Library preparation and sequencing

For library generation for each individual sample, equimolar amounts of the 18 different LD-PCR amplicons were pooled and purified using Agencourt AMPure XP Beads (Beckman Colter, Krefeld, Germany) according to manufacturer's instructions. The DNA concentration of the purified amplicon pools was measured using the Qubit dsDNA HS Assay kit on the Qubit 2.0 fluorometer (Life Technologies, Darmstadt, Germany) and diluted to 0.2 ng/μl. Indexed paired-end libraries were generated from 5 μl of diluted amplicon pools using the Nextera XT DNA Sample Preparation Kit (Illumina, Munich, Germany) according to manufacturer's protocol. Successful tagmentation was validated by ascertainment of fragment length of the libraries using the High Sensitivity DNA kit on the 2100 Bioanalyzer (Agilent Technologies, Hamburg, Germany). The final library (1.8 pM) was subjected to a 300 cycle sequencing run using the MiniSeq System Mid-Output Kit on the MiniSeq instrument (Illumina).

2.5 | Bioinformatic analysis

The data analysis was performed using the megSAP pipeline (https://github.com/imgag/megSAP) developed at the Institute of Medical Genetics and Applied Genomics, University Hospital of Tübingen, (Tübingen, Germany). Here, we briefly list the main tools of the pipeline: SeqPurge (v. 0.1-935; Sturm, Schroeder, & Bauer, 2016) was used for adapter and quality trimming, BWA mem (v. 0.7.17; Li, 2013) for read mapping, samblaster (v. 0.1.24; Faust & Hall, 2014) for duplicate removal, ABRA2 (v. 2.18; Mose, Perou, & Parker, 2019) for indel-realignment, freebayes (v. 1.2.0; Garrison & Marth, 2012) for variant calling, and Ensembl VEP (v. 94.5; McLaren et al., 2016) for variant annotation. Several tools from the ngs-bits toolset (https://github.com/imgag/ngs-bits) were used for quality control, annotation of in-house variant frequencies, copy-number variant calling, and other minor processing steps. MultiQC (v. 1.7; Ewels, Magnusson, Lundin, & Käller, 2016) was used to generate the quality metrics summary for all samples. The whole bioinformatics analysis and all genomic coordinates given in this manuscript are based on the GRCh37 genome. Variant designation is based on the NCBI Reference Sequence NM_019098.4.

2.6 | In silico analysis of intronic variants

Assessment of splicing changes was performed with the Alamut Genova software v1.4 (http://www.interactive-biosoftware.com) using default parameters. In addition, we used the Human Splicing Finder (HSF) 3.1 tool (www.umd.be/HSF/; Desmet et al., 2009) with default parameters which enables the detection of auxiliary motifs through the incorporation of all available matrices for splicing enhancers and silencers.

2.7 | Splicing analysis using minigene assays

For each candidate cryptic splice site variant, a DNA fragment including the variant and its flanking regions was amplified from selected patients using a proofreading polymerase and standard PCR protocols. Since all variants were present in heterozygous state in the respective patients, both the normal and the variant allele could be coamplified. Cloning into the exon-trapping vector pSPL3, transfection of human embryonic kidney (HEK) 293T cells,

RNA isolation and reverse transcriptase polymerase chain reaction (RT-PCR) were performed as described previously (Weisschuh, Wissinger, & Gramer, 2012).

2.8 | Relative quantification of RT-PCR products

RT-PCR from cDNA obtained from HEK 293 T cells (ATCC, Teddington, UK) transfected with plasmid constructs harboring the mutant c.1663-1205A-allele was performed with a 5′ FAM (6-carboxyfluorescein) labeled forward primer located in exon 14 (5-CCGTTCTCTATTTGCCTGGT-3′) and a reverse primer located in the pSPL3 tat2 exon (5-GATCCATTCGACCAATTCACT-3′) using the AmpliTaqGold polymerase (Thermo Fisher Scientific, Karlsruhe, Germany). FAM-labeled RT-PCR products were diluted 1:100 in water, mixed with 1 μl of GeneScan ROX500 size standard (Life Technologies, Darmstadt, Germany) and 8 μl of Hi-Di Formamide (Life Technologies) in a total volume of 10 μl. Mixes were separated by capillary electrophoresis on an ABI 3130XL Genetic Analyzer instrument (Life Technologies). The area-under-the-curve (AUC) was calculated with GeneMapper 5 (Life Technologies) software. Ratios of splicing products were determined as the AUC for individual peaks divided by the sum of AUC of all differentially spliced products.

2.9 | Haplotype analysis

Three annotated microsatellites in the vicinity of *CNGB3* (D8S1707, D8S167, and D8S1119) were analyzed as described previously (Mayer et al., 2017).

3 | RESULTS

3.1 | NGS quality metrics

NGS performed on an Illumina Miniseq instrument achieved on average 788,000 reads per sample with 340,000 being the lowest and 1,300,000 being the highest number of reads obtained. One hundred megabases (Mb) was the average amount of usable sequence bases obtained and 98.0% were mapped on target. The latter value corresponds to the region that could be amplified using the LD-PCR approach. As outlined in the Section Material and Methods, 10% of the target region could not be amplified despite several attempts evaluating different primer sequences, polymerases, and PCR conditions. More specifically, three intronic regions failed to be amplified. The biggest gap of 10,290 bp is located in intron 3 (GRCh37 chr 8:877,002,41-877,105,31). Two smaller gaps of 4,021 bp and 3,766 bp are located in intron 10 (GRCh37 chr 8:876,477,96–876,518,17) and intron 13 (GRCh37 chr 8:876,257,41-876,295,07), respectively. The overall mean coverage was 577 ± 116 with 89.2% of target sequence covered with at least 50 reads.

3.2 | Detection rate of known variants

In addition to the 33 monoallelic *CNGB3* cases, we sequenced six patients who were known to be homozygous for the recurrent c.1148del variant in *CNGB3*. These patients were mainly sequenced for haplotyping purposes in a different research context (data not shown). However, they also served as positive controls with respect to the detection rate and were used for subsequent filtering of candidate variants. Using our data analysis pipeline, we were able to redetect all previously known variants in the six patients homozygous for the

c.1148del variant and in the 33 single heterozygous *CNGB3* cases. This results in a detection rate of 100% for the previously known variants. The variants of the 33 single heterozygous *CNGB3* cases are listed in Table S2. For the sake of brevity, we will refer, in the following, to the known first alleles as coding variants.

3.3 | Filtering of variants

Whole-gene targeted sequencing of the *CNGB3* gene led to the identification of a mean of 257 variants per patient after exclusion of the known coding variants. After filtering out variants with a MAF greater than or equal to 0.01 in 1,000 Genomes and gnomAD population frequency databases, an average of 17 rare intronic variants per sample remained. Five patients were found to harbor a second pathogenic variant in the coding region of *CNGB3* that had escaped the prior screening based on Sanger sequencing and/or denaturing high performance liquid chromatography (Table S3). Segregation analyses could be performed in three of these cases and confirmed compound heterozygosity. In the remaining 28 patients a second filtering step was performed, in which all intronic variants were excluded that were found *in cis* with pathogenic variants in the solved cases (i.e., five bi-allelic mutation carriers and six patients homozygous for the c.1148del variant). The rationale behind this filtering step was that we considered it unlikely that a patient harbored three pathogenic variants in *CNGB3*. Where possible, we also performed segregation analysis to filter out candidate intronic variants *in cis* to the known coding variant. Variants that merely affected homonucleotide stretches or short tandem repeats were discarded. After having performed all filtering steps, 47 variants remained (Table S4). The position of each variant relative to the nearest canonical splice site ranged between 12 and 18,721 nucleotides. Eleven variants were seen in more than one patient (ranging from two times to 14 times), whereas 36 variants were unique. All variants except one were seen in heterozygous state. In addition to the 47 intronic variants, we identified one variant located 1,368 nucleotides upstream from the start codon and one variant located 1,652 nucleotides downstream of the stop codon (data not shown). Whether these two variants exert a pathogenic effect based on expression regulation or transcript stability remains unknown and was not analyzed.

3.4 | In silico prioritization

In silico assessment of intronic variants using four splicing algorithms embedded in the Alamut Genova software is shown in Table S4. Because the *CNGB3* gene is not known to have alternative transcripts, we did not include scores that predicted the disruption or weakening of noncanonical splice sites. Twenty-six variants were not predicted to create novel splice sites by any algorithm. Eleven variants were predicted by one algorithm to create or strengthen a cryptic splice site, seven variants by two and one variant by three algorithms, respectively. Only two variants were predicted to create or strengthen a potential donor splice site by all four algorithms. Variants were selected for further analysis by means of heterologous splice assays only if they created cryptic splice sites with a relative strength of more than 75% of the respective maximal score in at least two out of four splice prediction algorithms. This was only the case for three variants, namely c.212-3599T>A, c.1663-1205G>A and c.1663-2137C>T (shown in bold in Table S4), which were all predicted to create or strengthen a cryptic donor splice site. The typical length of an exon in

humans is less than 200 bp (Sakharkar, Chow, & Kangueane, 2004). Potential complementary acceptor splice sites located not more than 200 bp upstream of the cryptic splice donor were identified for all three variants (Figure 1).

3.5 | Splicing analysis using minigene assays

We aimed to analyze the potential effect on splicing for the three candidate splice site variants c.212-3599T>A, c.1663-1205G>A and c.1663-2137C>T in more detail. Since there is no substantial *CNGB3* expression in blood cells or other accessible tissues, we made use of a heterologous splicing assay in HEK293T cells to test mutant and wildtype *CNGB3* minigene constructs in direct comparison. For the analysis of c.1663-1205G>A and c.1663-2137C>T, which are both located in intron 14, amplicons were generated that comprised not only intron 14 but also both neighboring exons plus 500 bp of flanking intronic sequence. The third candidate variant c.212-3599T>A is located in intron 3 which comprises more than 13 kb. It was not possible to clone the entire intron 3 plus flanking exons into pSPL3. Hence, we cloned an amplicon which comprised the cryptic donor site plus two putative complementary acceptor splice sites and putative branch points.

The c.212-3599T>A variant did not exert any visible effect in the splicing assay (Figure 2c). In contrast, HEK293T cells transfected with plasmid constructs harboring the mutant c.1663-1205A-allele yielded two RT–PCR products one of which was slightly larger than the single product derived from cultures transfected with the wildtype c.1663-1205G-allele. Subsequent sequencing of subcloned RT-PCR products showed that the major product derived from the transfection with the mutant c.1663-1205A-allele comprised not only exons 14 and 15, but also a pseudoexon of 34 nucleotides spliced between both canonical exons (Figure 2a). This pseudoexon had exactly been predicted in silico (Figure 1). The minor product corresponds to the correctly processed transcript. To quantify the splicing defect more precisely we performed RT-PCR with a FAM-labeled forward primer and separated the PCR products by capillary electrophoresis. From the AUC of fluorescence intensity for each fragment we calculated the relative abundance of correctly versus aberrantly spliced RT-PCR products (Figure S1) which is 32% versus 68%. The aberrant transcript would lead, if translated, to an insertion of 32 novel amino acids followed by a premature termination codon. Taking into account that the variant is leaky (i.e., produces residual amounts of correct transcript), the nomenclature of this variant is c.1663-1205G>A/p.(G555Lfs*33, =). The second variant located in intron 14, c.1663-2137C>T, also exerted a splicing defect in the minigene assays. HEK293T cells transfected with plasmid constructs harboring the mutant c.1663-2137T-allele yielded a single RT-PCR product which was clearly larger than the single product derived from cultures transfected with the wildtype c.1663-2137C-allele. Subsequent sequencing of the aberrant transcript revealed an inframe pseudoexon of 99 nucleotides spliced between exons 14 and 15 (Figure 2b). This spliced in pseudoexon is one nucleotide shorter than was predicted in silico (Figure 1). The aberrant transcript would lead, if translated, to an insertion of eight novel amino acids followed by a premature termination codon, hence the nomenclature of this variant is c.1663-2137C>T/p.(G555Afs*9). If the mutant transcripts from both variants are not targeted to nonsense-mediated decay (NMD), the translated proteins are considered to be nonfunctional since they will lack the cGMP binding site.

Variants c.1663-1205G>A and c.1663-2137C>T have been submitted to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/, Landrum et al., 2014) with accession codes SCV000926194 and SCV000926195, respectively.

3.6 | Cohort screening for the c.1663-1205 G>A and c.1663-2137C>T variants

Our entire cohort of patients with a primary clinical diagnosis of ACHM comprises 1100 index patients. Among these, approximately 240 are still unsolved. We screened these patients for the c.1663-1205G>A and c.1663-2137C>T variants by means of amplicon sequencing and identified four additional cases heterozygous for c.1663-1205G>A, all of which were known to harbor another pathogenic coding variant in *CNGB3*. Segregation analysis to prove *trans* configuration of the coding variant and the deep intronic variant was only possible in one patient. We did not identify additional cases harboring the c.1663-2137C>T variant. Table 1 shows the genotypes of all patients that harbor the c.1663-1205G>A and c.1663-2137C>T variants.

3.7 | Founder effect analysis of the c.1663-1205G>A variant

To study the origin of the c.1663-1205G>A variant, we performed genotyping of three microsatellites in the vicinity of *CNGB3*. Segregation analysis with these microsatellite markers revealed that within our patient cohort there is a mix of recurrent mutational events and to a lesser extent founder effects for the c.1663-1205G>A variant (Figure S2). Whether the c.1663-2137C>T variant is a founder mutation could not be established since segregation analysis was only feasible in one of the two carriers.

4 | DISCUSSION

To the best of our knowledge, this is the first report of deep intronic variants in *CNGB3* presumably leading to functional *null* alleles. The genetic evidence that the c.1663-1205G>A variant causes ACHM is compelling. While it is absent in 31,416 alleles present in the population database gnomAD, it was observed in 18 patients from this study which makes it the eighth most frequent *CNGB3* mutation in our comprehensive ACHM cohort (Table 2). Segregation analysis was not feasible for all patients but we could verify that the c.1663-1205G>A variant was in *trans* to the coding variant in ten patients. For one patient only one parental sample was available and genotyping was indicative of a *trans* configuration of variants. The c.1663-2137C>T variant is less frequent in our patient cohort as it was found only in two patients. *Trans* configuration could be established in one patient. It is a rare variant in the general population, being only present once in heterozygosis in 30,756 alleles sourced from gnomAD.

In addition to the genetic evidence, the c.1663-1205G>A and the c.1663-2137C>T variants demonstrated a distinct splicing defect in our minigene assays as they both cause the insertion of a pseudoexon into the transcript, leading to premature termination codons, respectively. In contrast to the c.1663-2137C>T variant, the splicing defect exerted by the c.1663-1205G>A variant is not fully penetrant, at least in HEK293T cells, which upon transfection with the mutant allele also showed the correctly spliced transcript, albeit at a considerable lower level than the aberrantly spliced product (30% wildtype vs. 70% mutant).

All patients who are heterozygous for the c.1663-1205G>A variant carry either a frameshift, startloss, nonsense or splice site variant on the second allele. We therefore conclude that the c.1663-1205G>A variant may represent a mild allele associated with reduced levels of wildtype *CNGB3* transcripts (and eventually CNGB3 protein) that is unmasked as being deleterious if occurring in compound heterozygous state with a loss-of-function allele.

Haplotyping with three microsatellite markers surrounding the *CNGB3* locus showed that seven different alleles (Figure S2) are associated with the c.1663-1205G>A variant, suggesting a mutation hotspot rather than a founder effect. However, shared alleles were found in two and three families, respectively, and are in favor of founder events piggybacking on independent mutational events.

HEK293T cells transfected with plasmid constructs harboring the wildtype and mutant alleles of the candidate intronic variant c.212-3599T>A did not show any splicing defect. The derived RT-PCR products comprised only the two internal exons of the pSPL3 vector (i.e., tat1 and tat2) without a pseudoexon spliced between (Figure 2c). Hence, any putative pseudoexon created by the c.212-3599T>A variant was not recognized by the splicing machinery of the HEK293T cells. Since it was not possible to clone the entire intron harboring the variant into the splicing vector, we generated a minigene construct that includes two putative complementary splice sites and branch points. Hence, putative auxiliary *cis*-acting factors located upstream or downstream of the cloned sequence might be lacking in the construct. Similarly, the HEK293T cells might lack retina-specific splicing proteins (i.e., *trans* splicing factors) that are necessary for the recognition of a pseudoexon created by the c.212-3599T>A variant. Pseudoexon insertion can be tissue specific, as was shown for the most frequent Leber congenital amaurosis-associated *CEP290* variant, c.2991+1655A>G (den Hollander et al., 2006; Dulla et al., 2018). Hence, variant c.212-3599T>A that showed no effect on splicing in our study may still be proven to result in aberrant splicing when assessed in a retinal cell line.

Of the patients analysed in this study, 12 remained unsolved after targeted whole gene sequencing of *CNGB3*. This might be due to a number of reasons. First of all, we cannot be sure that the phenotype in all patients is related to mutations in *CNGB3*. Inherited retinal disorders (IRDs) often present with considerable clinical overlap (Sahel, Marazova, & Audo, 2014) which can preclude the assessment of a diagnosis on the basis of the disease phenotype alone, especially if patients are very young and follow-up data (i.e., progression of the disease) are not available. Because of this clinical overlap, we have included patients not only with a clinical diagnosis of ACHM, but also with other related diagnoses such as cone dystrophy, Morbus Stargardt or macular dystrophy. However, it is of course possible that the initial diagnosis in these patients is correct. From the 12 unsolved patients five were not diagnosed with ACHM but with a related disease (Table S2). The presence of monoallelic variants in *CNGB3* in these patients can be incidental, especially for recurrent alleles such as p.(R403Q) or c.1148del, with minor allele frequencies of 0.4% and 0.17% in population datasets, respectively. In fact, in the further course of this study, from the four unsolved patients who are heterozygous carriers of the frequent p.(R403Q) allele three could be solved in other research projects with pathogenic variants in other genes. One patient (MST 137) was shown to harbor two pathogenic variants in *ABCA4*, demonstrating that (a)

the clinical diagnosis of Morbus Stargardt had been correct, and (b) that the presence of the p.(R403Q) allele in this patient was incidental. Likewise, one patient (RCD 246) who was diagnosed with cone rod dystrophy could be solved with pathogenic variants in *CDHR1*. Another patient (ZD 203) was shown to harbor two pathogenic variants in the *ATF6* gene that had escaped a prior screening. To confirm that the phenotype in all patients included in this study is indeed caused by *CNGB3*, a comprehensive IRD panel sequencing approach could potentially exclude putative disease-causing variants in other IRD genes. Yet, our cohort was only pre-screened for mutations in the six known ACHM genes within our ACHM-related research projects. Another possible explanation why we were not able to solve all patients is that some intronic variants that were discarded in the filtering process in fact do exert a splicing defect. It has been shown that pseodoexon insertion most often is due to the generation or strengthening of cryptic splice sites, but it can also be promoted by the creation of exonic splicing enhancers (ESEs) or the abolishment of exonic splicing silencers (ESSs; Vaz-Drago, Custódio, & Carmo-Fonseca, 2017). As we focused on cryptic splice sites only, we cannot exclude that some of the intronic variants we identified lead to pseudoexon inclusion based on the formation of ESEs or the repression of ESSs. And last, disease-causing variants might be located in a region of *CNGB3* that could not be sequenced. As has been outlined before, we could not amplify 10% of the target region due to the presence of numerous repeat elements including *Alu* elements, which in turn often contain sequence motifs that resemble splice sites (Lev-Maor et al., 2003; Vorechovsky, 2010).

Deep-intronic splice variants are usually not identified in routine diagnostics due to its focus on protein-coding regions. However, the use of novel approaches such as whole genome sequencing or targeted locus resequencing have facilitated the scanning of an entire gene and uncovered deep-intronic splice mutations in multiple IRD genes such as *ABCA4* (Bauwens et al., 2019; Bax et al., 2015; Braun et al., 2013; Sangermano et al., 2019; Schulz et al., 2017; Zaneveld et al., 2015; Zernant et al., 2014), *USH2A* (Liquori et al., 2016; Vaché et al., 2012), *CEP290* (den Hollander et al., 2006), and others.

This is the first report of pathogenic deep-intronic variants in *CNGB3*. Screening of our entire cohort showed that the c.1663-1205G>A variant is among the ten most frequent mutations in *CNGB3*. Our study demonstrates that sequencing the entire gene in combination with a stringent filtering process can aid in identifying the second pathogenic allele in patients harboring a single disease-causing variant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank all patients and referring clinicians for their participation in the study. In particular, the authors acknowledge the contribution of Bernhard Jurklies (Wuppertal, Germany) and Rosanna Pescini Gobert (Lausanne, Switzerland). The authors would also like to thank Bin Guan, Ehsan Ullah, and Cara Lwin for WGS analysis and Sanger confirmation of family CHRO 1205. This work was funded in parts by a grant of LABEX LIFESENSES [reference ANR-10-LABX-65] supported by French state funds managed by the Agence Nationale de la Recherche within the Investissements d'Avenir program [ANR-11-IDEX-0004-0] to Isabelle Audo and by National Eye

Institute intramural funds to Robert B. Hufnagel. Isabelle Meunier, Isabelle Audo, Katarina Stingl, and Susanne Kohl are members of the European Reference Network dedicated to Rare Eye Diseases (ERN-EYE).

Funding information

Agence Nationale de la Recherche, Grant/Award Number: LABEX LIFESENSES [reference ANR-10-LABX-65]

REFERENCES

Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, … De Baere E (2019). ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novelNovel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. Genetics in Medicine, 21(8), 1761–1771. [PubMed: 30670881]

Bax NM, Sangermano R, Roosing S, Thiadens AA, Hoefsloot LH, van den Born LI, … Cremers FP (2015). Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. Human Mutation, 36(1), 43–47. 10.1002/humu.22717 [PubMed: 25363634]

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, … Stone EM (2013). Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. Human Molecular Genetics, 22(25), 5136–5145. 10.1093/hmg/ddt367 [PubMed: 23918662]

Buratti E, Chivers MC, Kralovicova J, Romano M, Baralle M, Krainer AR, & Vorechovsky I (2007). Aberrant 50 splice sites in human disease genes: Mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Research, 35(13), 4250–4263. 10.1093/nar/gkm402 [PubMed: 17576681]

den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, … Cremers FP (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. American Journal of Human Genetics, 79(3), 556–561. 10.1086/507318 [PubMed: 16909394]

Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, & Béroud C (2009). Human splicing finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Research, 37(9), e67. 10.1093/nar/gkp215.e67.

Dulla K, Aguila M, Lane A, Jovanovic K, Parfitt DA, Schulkens I, … Cheetham ME (2018). Splice-modulating oligonucleotide QR-110 restores CEP290 mRNA and function in human c.2991+1655A>G LCA10 models. Molecular Therapy. Nucleic Acids, 12, 730–740. 10.1016/j.omtn.2018.07.010 [PubMed: 30114557]

Ewels P, Magnusson M, Lundin S, & Käller M (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048. 10.1093/bioinformatics/btw354 [PubMed: 27312411]

Faust GG, & Hall IM (2014). SAMBLASTER: Fast duplicate marking and structural variant read extraction. Bioinformatics, 30(17), 2503–2505. 10.1093/bioinformatics/btu314 [PubMed: 24812344]

Garrison E, & Marth G (2012). Haplotype-based variant detection from short-read sequencing. arXiv, 1207, 3907. [q-bio.GN].

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, & Maglott DR (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research, 42(Database issue), D980–D985. 10.1093/nar/gkt1113 [PubMed: 24234437]

Lev-Maor G, Sorek R, Shomron N, & Ast G (2003). The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. Science, 300(5623), 1288–1291. 10.1126/science.1082588 [PubMed: 12764196]

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303, 3997v1. [q-bio.GN].

Liquori A, Vaché C, Baux D, Blanchet C, Hamel C, Malcolm S, … Roux AF (2016). Whole USH2A gene sequencing identifies several new deep intronic mutations. Human Mutation, 37(2), 184–193. 10.1002/humu.22926 [PubMed: 26629787]

Mayer AK, Van Cauwenbergh C, Rother C, Baumann B, Reuter P, De Baere E, … ACHM Study Group (2017). CNGB3 mutation spectrum including copy number variations in 552 achromatopsia patients. Human Mutation, 38(11), 1579–1591. 10.1002/humu.23311 [PubMed: 28795510]

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, … Cunningham F (2016). The ensembl variant effect Predictor. Genome Biology, 17(1), 122. 10.1186/s13059-016-0974-4 [PubMed: 27268795]

Mose LE, Perou CM, & Parker JS (2019). Improved Indel detection in DNA and RNA via realignment with ABRA2. Bioinformatics, 35(17), 2966–2973. 10.1093/bioinformatics/btz033.Jan. [PubMed: 30649250]

Sahel JA, Marazova K, & Audo I (2014). Clinical characteristics and current therapies for inherited retinal degenerations. Cold Spring Harbor Perspectives in Medicine, 5(2), a017111. 10.1101/cshperspect.a017111

Sakharkar MK, Chow VT, & Kangueane P (2004). Distributions of exons and introns in the human genome. In Silico Biology, 4(4), 387–393. [PubMed: 15217358]

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, … Cremers FPM (2019). Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genetics in Medicine, 21(8), 1751–1760. 10.1038/s41436-018-0414-9. Jan. [PubMed: 30643219]

Schindelin J, Rueden CT, Hiner MC, & Eliceiri KW (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. Molecular Reproduction and Development, 82(7-8), 518–529. 10.1002/mrd.22489 [PubMed: 26153368]

Schulz HL, Grassmann F, Kellner U, Spital G, Rüther K, Jägle H, … 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. Investigative Ophthalmology & Visual Science, 58(1), 394–403. 10.1167/iovs.16-19936 [PubMed: 28118664]

Sturm M, Schroeder C, & Bauer P (2016). SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. BMC Bioinformatics, 17, 208. 10.1186/s12859-016-1069-7 [PubMed: 27161244]

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, & Leunissen JA (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research, 35(Web Server issue), W71–W74. 10.1093/nar/gkm306 [PubMed: 17485472]

Vaché C, Besnard T, le Berre P, García-García G, Baux D, Larrieu L, … Roux AF (2012). Usher syndrome type 2 caused by activation of an USH2A pseudoexon: Implications for diagnosis and therapy. Human Mutation, 33(1), 104–108. 10.1002/humu.21634 [PubMed: 22009552]

Van Schil K, Naessens S, Van de Sompele S, Carron M, Aslanidis A, Van Cauwenbergh C, … De Baere E (2018). Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. Genetics in Medicine, 20(2), 202–213. 10.1038/gim.2017.97 [PubMed: 28749477]

Vaz-Drago R, Custódio N, & Carmo-Fonseca M (2017). Deep intronic mutations and human disease. Human Genetics, 136(9), 1093–1111. 10.1007/s00439-017-1809-4 [PubMed: 28497172]

Vorechovsky I (2006). Aberrant 3′ splice sites in human disease genes: Mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Research, 34(16), 4630–4641. 10.1093/nar/gkl535 [PubMed: 16963498]

Vorechovsky I (2010). Transposable elements in disease-associated cryptic exons. Human Genetics, 127(2), 135–154. 10.1007/s00439-009-0752-4 [PubMed: 19823873]

Weisschuh N, Wissinger B, & Gramer E (2012). A splice site mutation in the PAX6 gene which induces exon skipping causes autosomal dominant inherited aniridia. Molecular Vision, 18, 751–757. [PubMed: 22509105]

Zaneveld J, Siddiqui S, Li H, Wang X, Wang H, Wang K, … Chen R (2015). Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. Genetics in Medicine, 17(4), 262–270. 10.1038/gim.2014.174 [PubMed: 25474345]

Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, … Allikmets R (2014). Analysis of the ABCA4 genomic locus in Stargardt disease. Human Molecular Genetics, 23(25), 6797–6806. 10.1093/hmg/ddu396 [PubMed: 25082829]

c.1663-1205G>A

atatttttaaaaataaaaaataaatcttttaa
tctcgagttgattttagtgtattctaattctt
tcaaAGtatggttaatgagaatatttgattaa
ctcaaataacccagtcccctcctaagccaagt
aagtgaatttattgtattaatgctatttttt
ttatAGctcaaggtgatcaaggcactgctcat
tcagacatGTaagtactactagccaaata

Splice site type	HSF matrices	MaxEnt	NNSplice
acceptor	77.88	4.43	-
acceptor	85.92	8.46	0.99
donor	89.13	8.31	0.98

c.1663-2137C>T

tagcattaatagtatgtaaataaatataagat
ccatctatattggaatttctacttttatctgt
ttttgagtataaatatatgcttaatattttgc
acAGgccaggtgcagtggctcacgactgtaat
tccaacactttgggaggccaaggtggcaggat
cacataagtccaggagttcaagacaagcctgg
acaacatgGTaaaaccctgtctctacaat

Splice site type	HSF matrices	MaxEnt	NNSplice
acceptor	89.28	4.76	0.45
donor	79.8	4.79	0.95

c.212-3599T>A

aaatttgaccaaagtagtaccatcccccacac
caaagattctaaccagtgaaggcattgtcaag
tcttgtgcttttcatatccatctaaaacaagc
caactgagtgtctttttaatttacaaccttat
tcccatattAGgttctttctttgaatcctcc
tgtccAGcatttatgtaatgtcactgataaaa
tgtggaagGTaaatggatttccacatcac

Splice site type	HSF matrices	MaxEnt	NNSplice
acceptor	82.98	7.27	0.68
acceptor	86.88	4.01	-
donor	86.88	8.88	0.94

FIGURE 1.

Cryptic donor sites strengthened by the deep intronic variants and complementary cryptic acceptor sites. The sequence snippets show cryptic donor sites (blue) which are strengthened by the respective intronic variant (red). Putative complementary acceptor sites calculated with three algorithms embedded in the Human Splicing Finder 3.1 are shown in green. Only those acceptor sites are indicated that were predicted with a relative strength of more than 75% of the respective maximal score in at least two out of the three algorithms tested. invariable GT and AG dinucleotides are shown in capital letters

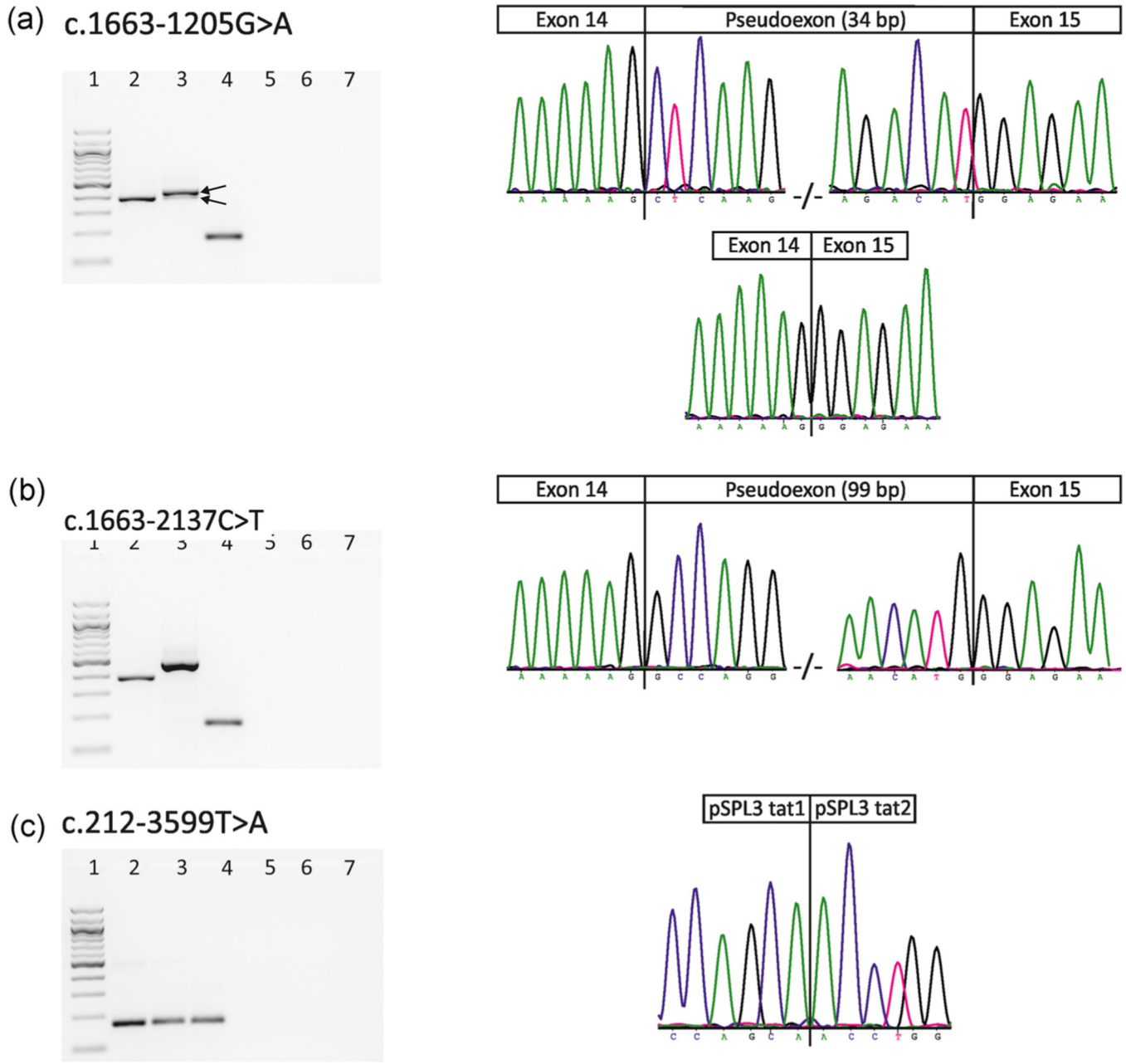

FIGURE 2.

Qualitative analysis of RT-PCR products from minigene splicing assays. Gel loading is as follows: Lane 1, 100 bp ladder size standard. RT-PCR products derived from HEK293T cells transfected with plasmid constructs harboring the wildtype allele of the respective variant (lane 2), while those with the mutant allele are shown in lane 3. Transfection with empty pSPL3 vector (lane 4) and untransfected cells (lane 5) served as controls. NRT (lane 6), no reverse transcriptase control; NTC (lane 7), no template control. (a) RT-PCR revealed one product in HEK293T cells transfected with plasmid constructs harboring the wildtype c.1663-1205G-allele and two products (indicated by arrows) in cells transfected with the mutant c.1663-1205A-allele. Sequence analysis of RT-PCR products shows that the major

RT-PCR product obtained from cells transfected with the mutant c.1663-1205A-allele comprises a pseudoexon spliced between exons 14 and 15 while the minor product represents correct splicing of exon 14 to exon 15. (b) Variant c.1663-2137C>T uniformly causes insertion of a pseudoexon of 99 nucleotides spliced between exons 14 and 15. (c) Variant c.212-3599T>A does not result in a splice defect. Sequence analysis of the single RT-PCR product obtained from cells transfected with the mutant c.212-3599A-allele shows only the two internal exons of the pSPL3 vector (i.e., tat1 and tat2) without a pseudoexon spliced between

TABLE 1

Patients carrying the c.1663-1205G>A or the c.1663-2137C>T variant

Patient ID	Allele 1	Allele 2	Segregation
CHRO 117	c.819_826del/p.(R274Vfs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 26	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 31	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 210	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 274	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 522	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 915	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 1205	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 1076	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 906	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 985	c.1148del/p.(T383Ifs*13)	c.1663-1205G>A/p.(G555Lfs*33)	Yes[a]
CHRO 1097	c.646C>T/p.(R216*)	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 771	c.991-3T>G/p.?	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 667	c.2T>C/p.?	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 273	c.1635A>T/p.Y545*	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 658	c.644-1G>C/p.?	c.1663-1205G>A/p.(G555Lfs*33)	Na
CHRO 983	c.1178+5G>T/p.?	c.1663-1205G>A/p.(G555Lfs*33)	Yes
CHRO 1138	c.819_826del/p.R274Vfs*13	c.1663-1205G>A/p.(G555Lfs*33)	Na
ZD 216	c.1148del/p.(T383Ifs*13)	c.1663-2137C>T/p.(G555Afs*9)	Na
CHRO 1069	c.1190_1192del/p.(C397del)	c.1663-2137C>T/p.(G555Afs*9)	Yes

Notes: Variant designation based on NCBI Reference Sequence NM_019098.4. na, not analyzed.

[a]Only one parental sample was available and indicative of a *trans* configuration of variants.

TABLE 2

Ten most frequent *CNGB3* alleles in the entire achromatopsia cohort (n = 1100)

Nucleotide (NM_019098.4)	CNGB3 protein (NP_061971.3)	Total number of alleles
c.1148del	p.(T383Ifs*13)	730
c.819_826del	p.(R274Vfs*13)	49
c.1578+1G>A	p.?	33
c.1006G>T	p.(E336*)	27
c.1208G>A	p.(R403Q)	27
c.991-3T>G	p.?	24
c.886_896delinsT	p.(T296Yfs*9)	21
c.1663-1205 G>A	p.(G555Lfs*33)	18
c.1063C>T	p.(R478*)	10
c.1432C>T	p.(S435F)	10

The variant shown in bold is one of the deep intronic variants identified in this study and is the eighth most frequent *CNGB3* variant in our cohort.