Long-range PCR-based NGS applications to diagnose Mendelian retinal diseases

Jordi Maggi[1], Samuel Koller[1], Luzy Bähr[1], Silke Feil[1], Fatma Kivrak Pfiffner[1], James V. M. Hanson[2], Alessandro Maspoli[1], Christina Gerth-Kahlert[2], Wolfgang Berger[1,3,4,*]

[1]Institute of Medical Molecular Genetics, University of Zurich, 8952 Schlieren, Switzerland

[2]Department of Ophthalmology, University Hospital Zurich and University of Zurich, 8091 Zurich, Switzerland

[3]Zurich Center for Integrative Human Physiology (ZIHP), University of Zurich, Zurich, Switzerland

[4]Neuroscience Center Zurich (ZNZ), University and ETH Zurich, Zurich, Switzerland

*Corresponding author: berger@medmolgen.uzh.ch, +41 44 556 33 50

Abstract

Purpose: To develop a flexible, cost-efficient next-generation sequencing (NGS) protocol for genetic testing.

Methods: Long-range polymerase chain reaction (LR PCR) amplicons of up to 20 kb in size were designed to amplify entire genomic regions for a panel (*n*=35) of loci associated with retinal diseases (RDs). Amplicons were pooled and sequenced by NGS. The analysis was applied to 227 probands diagnosed with RD: (A) 108 previously molecularly diagnosed, (B) 94 without previous genetic testing, (C) 25 undiagnosed after exome sequencing (ES).

Results: The method was validated with 100% sensitivity on cohort A. LR PCR-based sequencing revealed likely causative variant(s) in 51% and 24% of proband from cohorts B and C, respectively.

Breakpoints of 3 CNVs could be characterized. LR PCR libraries spike-in extended coverage data of ES. Read phasing confirmed compound heterozygosity in 5 probands.

Conclusion: The proposed sequencing protocol provided deep coverage of the entire gene including intronic and promoter regions. Our method can be used (i) as a first-tier assay to reduce genetic testing costs, ii) to elucidate missing heritability cases, iii) to characterize breakpoints of CNVs at the nucleotide level, iv) to extend WES data to non-coding regions by spiking-in LR libraries, and v) to help with phasing of candidate variants.

Keywords: genetic testing; inherited retinal dystrophies; sequencing; NGS; diagnostics; long-range PCR; CNV; *ABCA4*; *PRPH2*; *BEST1*; phasing; missing heritability

Introduction

Inherited retinal diseases (RDs) are a group of disorders affecting the retina and its function. They are characterized by high phenotypic variability and genetic heterogeneity, including different modes of inheritance.[1] These disorders may present as either isolated or syndromic, progressive or stationary. Also, RDs may affect the entire retina (panretinal) or may be restricted to the macula. Moreover, they are often classified based on the primarily affected photoreceptor type; panretinal forms include retinitis pigmentosa (RP), cone-rod dystrophy (CRD), cone dystrophy (COD), and others, whereas macular dystrophies (MDs) include Stargardt disease (STGD), Best disease (BEST), and others.[1,2]

Due to the multitude of loci associated with RDs (almost 300 loci, https://sph.uth.edu/RetNet), molecular diagnostics often relies on targeted enrichment and high throughput sequencing, either exome (ES)[3–5] or gene panels[4–8]. Studies have reported that diagnostic yield when using these standard methods is typically between 50 and 76%, depending, amongst other factors, on the specific clinical subtype being investigated.[5–11]

Missing heritability is defined as the portion of the phenotypic trait that could not (yet) be explained by genotype data.[12] Since inherited diseases are, by definition, characterized by complete or near-complete heritability, missing heritability in these diseases indicates cases in which genetic testing could not identify a molecular diagnosis.[12]

Recent efforts to reduce missing heritability in RDs (and improve diagnostic output) have focused on non-coding regions of the genome, and on the detection of

structural or copy number variants (CNVs).[13–16] These studies were performed using customized microarray probes[14], customized capturing probes[13,15], or genome sequencing[16].

Several studies have focused on the *ABCA4* locus.[13,15,17–20] Variants in this gene are the cause of STGD1 (MIM 248200).[21] Previous studies have reported that 15-40% of STGD patients remain without a molecular diagnosis after standard genetic testing.[13,15,17–20] These studies identified CNVs and several non-canonical splice site and deep-intronic variants that have been shown to affect splicing, findings that partially explain missing heritability in STGD[13,15,17–20].

Here, we present a flexible and cost-effective method to comprehensively sequence loci of interest. The method relies on high-throughput sequencing of pooled LR PCR fragments of up to 20 kb in size. We designed primers to sequence 35 genomic loci that have been associated with RDs, particularly MDs and X-linked RP, including *ABCA4*, *PRPH2*, *BEST1*, *CRB1*, *CNGB3*, *RP1*, and *RPGR.*

Materials & Methods

Patients and family members

Unrelated probands (*N*=227) were referred to us for genetic testing from different eye clinics with a clinical diagnosis and information about family history. Samples (*N*=166) from 66 families were available for segregation analysis. Written informed consent was obtained from all probands and family members included in this study, which was conducted in accordance with the 2013 Declaration of Helsinki.

The probands included in this study were assigned to one of three groups: (i) validation cohort (molecular diagnosis previously established, *n*=108), (ii) missing

heritability cohort (no molecular diagnosis established with previous ES analysis, *n*=25), (iii) probands without previous genetic testing (*n*=94).

Genomic DNA

The majority of genomic DNA (gDNA) samples (*n*=204) were extracted in duplicate with the automated chemagic MSM I system according to the manufacturer's specifications (PerkinElmer Chemagen Technologie GmbH, Baesweiler, Germany). The remaining gDNA samples (*n*=23) were extracted at external labs and sent to us. Genomic DNA integrity and concentration were evaluated on a Nanodrop instrument (Life Technologies, Carlsbad, USA). Aliquots of gDNA were diluted to 10 ng/µl with ddH_2O for use in PCRs.

LR PCR primers

Primer for LR PCR were designed on NCBI's Primer-Blast to be 28 bp (range 26-30) long and to have a melting temperature (Tm) of 67° C (range 65°-68°C) under default settings[22]. The size of PCR products depended on the target locus, and ranged from 4.3 to 19.9 kb (mean size = 15.8 kb, median size = 17.6 kb). Primer for a total of 124 PCRs to amplify the following RD-associated loci were validated (synthesized by Microsynth AG, Balgach, Switzerland): *PPT1*, *ABCA4*, *CRB1*, *PCARE* (*C2orf71*), *EFEMP1*, *IMPG2*, *PROM1*, *MFSD8*, *CTNNA1*, *GUCA1A*, *GUCA1B*, *PRPH2*, *IMPG1*, *ELOVL4*, *DHS6S1*, *RP1L1*, *RP1*, *CNGB3*, *KCNV2*, *ATOH7*, *PDE6C*, *BEST1*, *C1QTNF5*, *PDE6H*, *RDH5*, *OTX2*, *NR2E3*, *RLBP1*, *GUCY2D*, *FSCN2*, *RAX2*, *TIMP3*, *RS1*, *RPGR*, *RP2*. The validated primers are listed on Table S1.

LR PCR

The LR PCR products used in this study were generated using Takara's *LA Taq* polymerases (Takara Bio, Kusatsu, Japan) according to the following PCR mixture: 30 µl total reaction volume, 11.5 µl of ddH_2O, 50 ng of template gDNA, 3 µl of 10x LA PCR Buffer II (Mg^{2+} plus), 3 µl of 10x Solution S (Solis BioDyne, Tartu, Estonia), 4.8 µl of dNTPs Mixture (2.5 mM each), 1.2 µl of each primer (10 mM), and 0.3 µl of *Takara LA Taq* (5 units/µl). Reactions were performed on a Veriti thermal cycler (Applied Biosystems, Foster City, USA) according to 2-step PCR conditions: 94°C for 2 min; 35 cycles of 98°C for 10 s and 68°C for 12 min; followed by 72°C for 10 min; and hold at 4°C.

PCR quality control and pooling

The expected size of amplicons was confirmed by electrophoresis on 0.6% agarose gels (run at 60 V) and 1 µl of the reaction mixture was used to measure the amplicon concentration with the Qubit dsDNA High Sensitivity Assay kit (ThermoFisher, Waltham, USA).

All amplicons that would be sequenced with the same index sequence (either all PCRs of a specific proband or non-overlapping amplicons of different probands) were diluted to 10 ng/µl in IDTE buffer (Integrated DNA Technologies, Coralville, USA). The volume of each diluted PCR added to the pool was proportional to its size, for a final pool of at least 130 µl in total volume.

Library construction and sequencing

Each pool was sheared on a Covaris M220 (Covaris, Woburn, USA) in a 130 µl Covaris AFA microtube (Covaris, Woburn, USA) to a target size of ~ 350-400 bp,

with the following settings: 50 W peak incident power, 20% duty factor, 200 cycles per burst, 65 s treatment time.

Successful fragmentation was checked by running 1 µl of the sheared pool with a Bioanalyzer High Sensitivity DNA kit on a Bioanalyzer 2100 instrument (Agilent Technologies, Santa Clara, USA).

The validated pools were then processed with the TruSeq DNA Nano Low Throughput Library Prep kit and TruSeq DNA Single Index Set A and B, according to the manufacturer's protocol (Illumina, San Diego, USA).

Molarities of the libraries were calculated and diluted to a 4 nM working concentration. The different libraries that were to be sequenced together (with different indexes) were pooled proportionally to their total genomic target size. Subsequently, these pools were denatured with 0.2N NaOH and loaded into a MiSeq Reagent Kit V2 (300-cycles) cartridge, according to the manufacturer's recommendations (Illumina, San Diego, USA). Paired-end sequencing (2x 151 cycles) was performed on a MiSeq instrument (Illumina, San Diego, USA).

A detailed step-by-step protocol is provided in the Supplementary Materials (Text S1).

ABCA4 capture sequencing

Custom capture probes (xGen Lockdown Probe Pools) were designed for the *ABCA4* locus by IDT (Integrated DNA Technologies, Coralville, USA). Genomic DNA libraries were constructed according to the ThruPLEX DNA-seq 96D kit protocol (Takara Bio, Kusatsu, Japan). Libraries were pooled equimolarly for a total of 1600 ng of DNA. Subsequently, the *ABCA4* capture probes were used according to the manufacturer's protocol to enrich the *ABCA4* locus (Integrated DNA Technologies,

Coralville, USA). The resulting libraries were sequenced on a MiSeq according to manufacturer's recommendations (Illumina, San Diego, USA).

Exome sequencing (ES)

ES was performed for 152 probands; 23 of them underwent ES in an external facility on an Illumina HiSeq instrument (AtlasBiolabs, Berlin, Germany) or on a SOLiD 5500xl system (CeGat, Tübingen, Germany), whilst 129 probands were sequenced in-house on a NextSeq 550 (Illumina, San Diego, USA). Of the 129 in-house ES, 31 were captured using the Nextera Rapid Capture Exome kit (Illumina, San Diego, USA), 65 with the TruSeq Exome kit (Illumina, San Diego, USA), and 33 with the xGen Exome Research Panel (Integrated DNA Technologies, Coralville, USA), according to the manufacturers' protocols.

Exome spike-in applications

To assess the feasibility of spike-in to ES experiments to enhance coverage of regions of interest, we tested two alternatives: by adding the custom xGen Lockdown Probe pool for the *ABCA4* locus to the TruSeq Exome kit capturing probes mix before hybridization, or by adding the processed LR PCR libraries just before final library denaturation.

An aliquot of the xGen Lockdown Probes was diluted to a concentration of 85.5 aM and a total of 250 amol were added to the TruSeq Exome capture mix for spike-in. Conversely, for the LR libraries spike-in, the amount of library to be added was calculated based on the ratio of the total LR library target regions and the ES target region.

Sequencing data analysis

Illumina sequencing data was aligned to the human reference genome hg19 with BWA, and variant calling was performed by GATK.[23,24] The resulting Variant Call Format files (VCFs) were annotated with Alamut Batch v1.8 (SophiaGenetics, Lausanne, Switzerland) using a gene list corresponding either to the loci present in the RetNet database (*n*=276) for ES data, or to the target loci for LR PCRs data. CNVs analysis on target-capture sequencing data (*ABCA4* capture and ES results) was performed using panelcn.mops[25] and the SeqNext module of SeqPilot v5.2 (JSI medical systems, Ettenheim, Germany).

LR PCR sequencing data of *RPGR*'s ORF15 was assembled *de novo* using SPAdes and the resulting contig was aligned to the reference sequence, as previously described.[26,27]

CNV validation and breakpoints characterization

Multiplex ligation-dependent probe amplification (MLPA; MRC Holland, Amsterdam, The Netherlands) was used to validate candidate CNVs identified by panelcn.mops and/or SeqNext in *ABCA4*, *BEST1*, *GUCY2D*, and *RPGRIP1* and to screen for CNVs in *ABCA4*, *BEST1*, and *PRPH2*.

Deletions encompassing exons 20 through 22 of *ABCA4*, exon 2 of *RS1*, exons 2 through 4 of *RP1*, and the entire *KCNV2* gene were identified by panelcn.mops and SeqNext or by comparison of coverage plots. These deletions were verified and characterized by LR PCR. The primers used are listed in Table S2. Briefly, primers were designed based on the estimated breakpoints location, LR PCR was performed, and the amplicon was sequenced as described above (see Amplicons pool library construction and sequencing). In addition, breakpoints were confirmed by Sanger sequencing (primers in Table S2).

To narrow down the breakpoint region of the *KCNV2* deletion, an Infinium CytoSNP-850K BeadChip array analysis was performed according to the manufacturer's protocol (Illumina, San Diego, USA).

The *ABCA4* deletion was further characterized and refined by targeted locus amplification (TLA; Cergentis, Utrecht, The Netherlands).

Chromosomal phasing

Segregation analysis was performed if samples from family members were available, either by Sanger sequencing (primers on Table S3) or by LR PCR sequencing as described above.

Alternatively, when no family members were available for testing, read phasing was performed. For this, reads in the BAM file were visually inspected for informative heterozygous variants between the two putative compound heterozygous candidate variants.

Results

The method was established for 35 loci associated with RDs. The total genomic target equals 1.81 Mb, and the longest locus is *CRB1*, with a total genomic target of 212 kb (Table 1). The entire regions of most target genes could be amplified and sequenced. Several intronic regions could not be amplified after several attempts: *CRB1* (3.8 kb in size, NC_000001.10:g.197420721-197424549), *EFEMP1* (0.5 kb, NC_000002.11:56102261-56102747), *IMPG1* (1.9 kb, NC_000006.11:76729633-76731567 and 1.6 kb, NC_000006.11:76633571-76635199), *CNGB3* (7.1 kb, NC_000008.10:g.87647022-87654129 and 3.8 kb, NC_000008.10:87625835-87629598), *TIMP3* (1.5 kb, NC_000022.10: 33202359-33203885).

Table 1. **Established long-range PCRs for retinal diseases-associated loci**. The first and second columns report the corresponding transcript identifiers, respectively, of the loci included in the panel, ordered by chromosome coordinates. The third column highlights the number of PCRs necessary to cover the locus comprehensively, and the fourth column shows the total size of genomic target sequence.

Locus	Transcript	# of PCRs needed	Genomic target size (bp)
PPT1	NM_000310.3	2	26’746
ABCA4	NM_000350.2	8	137’748
CRB1	NM_201253.2	14	212’898
PCARE	NM_001029883.2	1	18’628
EFEMP1	NM_001039348.2	5	60’569
IMPG2	NM_016247.3	6	94’768
PROM1	NM_006017.2	7	117’712
MFSD8	NM_152778.2	3	49’747
CTNNA1	NM_001323982.1	13	182’694
GUCA1A	NM_000409.4	2	27’191
GUCA1B	NM_002098.5	1	13’548
PRPH2	NM_000322.4	2	32’242
IMPG1	NM_001563.3	11	144’700
ELOVL4	NM_022726.3	2	34’792
DHS6S1	NA	1	13’473
RP1L1	NM_178857.5	3	51’877
RP1	NM_006269.1	1	16’930
CNGB3	NM_019098.4	12	172’052
KCNV2	NM_133497.3	1	15’214
ATOH7	NM_145178.3	1	9’275
PDE6C	NM_006204.3	3	53’794
BEST1	NM_004183.3	1	17’272
C1QTNF5	NM_015645.4	1	10’536
PDE6H	NM_006205.2	1	11’084
RDH5	NM_002905.3	1	6’122
OTX2	NM_021728.3	1	12’758
NR2E3	NM_014249.3	1	11’097
RLBP1	NM_000326.4	1	13’667
GUCY2D	NM_000180.3	1	18’939
FSCN2	NM_001077182.2	1	11’574
RAX2	NM_001319074.1	1	5’845
TIMP3	NM_000362.4	5	67’146
RS1	NM_000330.3	3	34’731
RPGR	NM_001034853.1	4	60’808
RP2	NM_006915.2	3	46’521

The primers for loci that require multiple PCRs were designed in a way that amplicons of adjacent regions overlap with each other. As an example, Figure 1

shows the coverage obtained for the *ABCA4* locus by different NGS assays: ES, custom capture probes, and LR PCRs.

As per design, ES only provides coverage for the coding portions of the locus. The custom capture probes result in coverage for the majority of the locus, including introns. However, gaps accounting for 15-20% of the locus are present. Finally, the LR PCR method can provide uniform coverage over the entire locus. Overlaps between neighboring LR amplicons are revealed by an increase in coverage. Similar results were obtained for the other loci included in the panel (Figure S1).

Figure 1. *ABCA4* coverage comparison of different assays. Screenshot of the Alamut Visual software showing the coverage results for the *ABCA4* locus from different next-generation sequencing assays. On the top, the software illustrates the relative exon locations. Below the gene structure, coverage plots for a typical exome sequencing assay (top), custom capture probes assay (middle) and, finally, the long-range polymerase chain reaction method (bottom).

In order to verify sensitivity of the method, 108 RD patients were selected for whom a molecular diagnosis had been previously established, corresponding to variants in a locus included in the panel. LR PCRs for the respective loci were performed and sequenced. The sequencing of the validation cohort resulted in 100% sensitivity to the previously identified putative pathogenic variants (Table S4). We found no evidence for allele dropout (ADO).

We envisioned the method to be useful in the following scenarios (i-v):

I. First-tier assay for probands without previous genetic testing

Candidate loci were sequenced for 94 non-syndromic RD cases composed of probands diagnosed with MD (92/94), CRD (1/94), and autosomal dominant COD (1/94) that had not previously undergone genetic testing. Locus selection was based on the provided clinical diagnosis and family history. This strategy resulted in the identification of a molecular diagnosis in 48/94 probands (51.1%) (Table S5).

If candidate loci sequencing did not reveal a molecular diagnosis, ES was performed. Second-tier ES analysis revealed likely pathogenic variants in 15 additional probands, resulting in a total of 63/94 molecularly diagnosed probands (Table S5). Thirty-one probands (33.0%) were still lacking a genetic diagnosis after first-tier LR PCR-based candidate loci sequencing and second-tier ES. A list of rare variants (gnomAD MAF < 1%) found in these samples is provided (Table S7).

II. Tackling missing heritability

The missing heritability cohort included 25 probands in whom ES was previously performed and a single likely pathogenic variant in a recessively-inherited locus was identified. LR PCRs of the respective loci were performed to assess the presence of a second likely pathogenic variant.

This strategy led to the discovery of additional likely pathogenic variants in *ABCA4* in 6/25 probands (23.1%). The most common likely causative variant identified is NM_000350.2:c.5603A>T (3 alleles), which is often excluded from analysis due to the relatively high minor allele frequency in gnomAD (6.65% in Non-Finnish European).[15] The remaining cases could be explained by the discovery of previously published deep intronic variants: NM_000350.2:c.4253+43G>A[18,19], NM_000350.2:c.4539+2064C>T[15,17], and NM_000350.2:c.5196+1056A>G[13,17]

(Table S6). Moreover, this strategy led to the detection of several deep-intronic variants of unknown significance, which require further analyses for interpretation of pathogenicity (Table S7).

III. CNV characterization

CNV analysis on capture data allowed for the identification of likely pathogenic exon-spanning deletions in *ABCA4*, *KCNV2*, *RP1*, and *RS1*.

A heterozygous *ABCA4* deletion spanning exons 20 through 22 and parts of the adjacent introns was identified in two unrelated STGD patients and confirmed by MLPA (Figure S2 and Table S5). A LR PCR product generated by using forward and reverse primers upstream of exon 19 and downstream of exon 23 (Table S2), respectively, was sequenced on a MiSeq instrument (Illumina, San Diego, USA). Results highlighted similar breakpoints as in a recently published recurrent deletion[15]; however, mismatches and a longer polyadenine stretch seem to be present around the breakpoints. The breakpoint region was also sequenced by TLA (Cergentis, Utrecht, The Netherlands), which appeared to confirm the 3' breakpoint at NC_000001.10:g.94511701. The 5' breakpoint probably lies within an A-rich sequence at NC_000001.10:g.94507655-94507699. Interestingly, the breakpoints are flanked by an AluY and an AluSx element on the 5'-(NC_000001.10:g.94507370-94507684) and 3'-side (NC_000001.10:g.94511717-94512041), respectively. Finally, a 10 bp microhomology sequence directly follows the 5'-side AluY element (NC_000001.10:g.94507690-94507699) and precedes the 3'-side AluSx element (NC_000001.10:g.94511701-94511710). However, unambiguous breakpoint identification was not possible due to repetitive flanking sequences (Figure S2 and Table 2).

A heterozygous deletion spanning the entire *KCNV2* locus was identified in a patient diagnosed with COD (Table S4). To narrow down the breakpoint region of the deletion, microarray analysis was performed with an 850K SNP chip. Based on these results, primers were designed for LR PCR (Table S2) and the resulting amplicon was sequenced. Sequence alignment revealed a deletion of 70'036 bp in size and a 3-bp microhomology sequence flanking the breakpoints (Table 2, Figure S3).
Finally, a heterozygous deletion involving exon 2 of *RS1* and a homozygous deletion spanning exons 2-4 of *RP1* were identified in a X-linked retinoschisis (XLRS) patient and a RP patient, respectively. The breakpoints of these deletions could be determined directly with the validated LR PCRs (Table S1). The *RP1* deletion is 11'116 bp in length with a 3-bp microhomology region flanking the breakpoints and has been described previously[28] (Table S4). The *RS1* deletion is 1'005 bp in length and features a 4-bp microhomology around the breakpoints (Table 1, Table S4 and Figure S4).

IV. Exome spike-in

In order to include the intronic regions of *ABCA4*, the IDT custom capture probes were added to the TruSeq Exome kit capture probes before the hybridization step. The resulting data provided low coverage of the intronic regions of the gene (Figure S5). However, achieving the appropriate dilutions and proportions to ensure that the *ABCA4* capture probes do not outnumber ES probes was challenging.
Similarly, LR PCR libraries spike-in was performed for the *ABCA4*, *RPGR*, and *PCARE* (*C2orf71*), to either increase coverage of poorly captured regions, such as *RPGR*'s ORF15 and *PCARE* exon 1, or to cover non-captured regions, such as introns. The method provided reliable data for variant detection over the entire loci,

including ORF15 (Figure S5). Moreover, it proved to be technically less challenging compared to custom probes spike-in. Finally, the extra target regions can be personalized for each individual sample.

V. Read phasing

In case of compound heterozygosity for recessive variants, chromosomal phasing for the candidate variants should be performed. Segregation analysis can achieve this, but only if family members are available for testing. Additional family members of 66 probands were available for segregation analysis in this study (66/227, 29.1%). Being able to perform chromosome phasing directly on an individual's sample would greatly help with interpretation of genetic testing results. By visually inspecting reads resulting from LR PCR sequencing (read phasing), it was possible to ascertain compound heterozygosity in 5 cases (Table S4-5). Figure S6 shows a graphic representation of the concept, along with the simplest specific example (patient ID S130, Table S4).

MD cohort

MD cases were present in all three groups analyzed in this study; 70 in the validation cohort, 13 in the missing heritability group, and 85 in the cohort in whom genetic testing had not been previously performed. The MD cohort (*N*=168) was composed of probands with clinical diagnosis of STGD (*n*=95), unspecified MD (*n*=45), Best disease (*n*=11), malattia leventinese (ML; *n*=9), XLRS (*n*=6), Sorsby fundus dystrophy (SFD; *n*=1), and adult vitelliform macular dystrophy (AVMD; *n*=1) (Table 2). The overall diagnostic yield equals 78.6%, with the STGD subcohort reaching 88.4% detection rate (Table 2 and Figure 2). Moreover, similar to previously reported

findings[11], *ABCA4* (*n*=75), *PRPH2* (*n*=17), and *BEST1* (*n*=10) are the most common contributors in our cohort (Figure 2).

Table 2. Macular dystrophies cohort summary.

Category	Total #	Total diagnosed	Total undiagnosed	Mean age at referral (years)	Diagnostic yield (%)
Overall	168	132	36	37.86	78.6
Stratification					
Validation	70	70	0	35.60	100.0
Missing heritability	13	4	9	35.08	30.8
No previous testing	85	58	27	40.14	68.2
Clinical diagnosis					
STGD	95	84	11	35.88	88.4
Unspecified MD	45	28	17	40.11	62.2
BEST	11	9	2	35.73	81.8
ML	9	5	4	47.89	55.6
XLRS	6	6	0	32.00	100.0
SFD	1	0	1	63.00	0.0
AVMD	1	0	1	67.00	0.0

Figure 2. Diagnostic yields and main contributing loci for the macular dystrophies cohort and for the Stargardt disease subcohort. Nested pie charts depicting diagnostic yield (inner ring) and the identified contributing loci (outer ring) for macular dystrophies (top) and Stargardt disease (bottom). The loci included in the long-range panel are shown in different shades of blue, whilst other retinal diseases-associated loci not included in the panel are shown in green. *ABCA4* and *PRPH2* are the main contributors in both cohorts.

A total of 161 unique, likely pathogenic variants were deemed disease-relevant in our cohort with 39 being novel (not found in Human Gene Mutation Database; summarized on Table 2). Of these 161 variants, 127 were SNVs (26 novel), 30 small

indels (10 novel), and 4 larger deletions (3 novel). The most common causal variants were NM_000350.2(*ABCA4*):c.5882G>A (35 alleles), NM_000322.4(*PRPH2*):c.514C>T (11 alleles), NM_000350.2(*ABCA4*):c.5603A>T (not in *cis* with other known causal variants, 10 alleles), NM_000350.2(*ABCA4*):c.5714+5G>A (9 alleles), NM_000350.2(*ABCA4*):c.[5461-10T>C;5603A>T] (7 alleles), NM_000350.2(*ABCA4*):c.[1622T>C;3113C>T] (6 alleles), and NM_000350.2(*ABCA4*):c.[2588G>C;5603A>T] (6 alleles). Among the novel findings (Table 2), there are three recurrent likely pathogenic variants: NM_000350.2(*ABCA4*):c.4958G>A (2 alleles), NM_152778.2(*MFSD8*):c.670A>T (3 alleles), and NM_133497.3(*KCNV2*):c.1096del (2 alleles, found once in an undiagnosed case).

Table 3. Novel likely pathogenic variants

Locus	cNomen	pNomen	gnomAD overall (%)	gnomAD max. (%)	ACMG class	CADD score
ABCA4	NM_000350.2:c.6731T>A	p.Val2244Glu	0	0	3	27.3
ABCA4	NM_000350.2:c.6428T>A	p.Met2143Lys	0	0	4	32.0
ABCA4	NM_000350.2:c.6323_6331delinsGGC	p.Met2108_Asn2111delinsArgHis	0	0	4	35
ABCA4	NM_000350.2:c.5924G>T	p.Gly1975Val	0	0	4	28.8
ABCA4	NM_000350.2:c.5690_5704del	p.Gln1897_Phe1901del	0	0	3	22.2
ABCA4	NM_000350.2:c.5691G>T	p.Gln1897His	0	0	3	23.8
ABCA4	NM_000350.2:c.5461-6T>C	p.?	0	0	3	14.93
ABCA4 [a]	NM_000350.2:c.4958G>A	p.Gly1653Glu	0	0	3	28.1
ABCA4	NM_000350.2:c.4609del	p.Thr1537Argfs*6	0	0	4	33
ABCA4	NM_000350.2:c.4383G>C	p.Trp1461Cys	0.0004	0.003	3	32
ABCA4	NM_000350.2:c.3323del	p.Arg1108Profs*40	0	0	5	34
ABCA4	NM_000350.2:c.3179A>C	p.Gln1060Pro	0.0008	0.0065	4	23.7
ABCA4 [b]	NM_000350.2:c.(2918+765_2918+775)_(3328+618_3328+662)del	p.Leu973_Asp2273delinsPheMetAlaArgValGluArgSerLeuGlyAsn	0	0	5	
ABCA4	NM_000350.2:c.1742C>A	p.Thr581Asn	0	0	3	26.7
ABCA4	NM_000350.2:c.1621_1622del	p.Leu541Thrfs*14	0	0	4	32
ABCA4	NM_000350.2:c.727_728dup	p.Tyr245Cysfs*18	0	0	5	26.2
ABCA4	NM_000350.2:c.676C>A	p.Arg226Ser	0.0068	0.0163	3	14.13
CRB1	NM_201253.2:c.1472A>T	p.Asp491Val	0	0	3	15.86
CRB1	NM_201253.2:c.2298G>A	p.Trp766*	0	0	5	36
OPA1	NM_130837.2:c.2987A>C	p.Lys996Thr	0	0	4	23.8
PROM1	NM_006017.2:c.2476G>C	p.Asp826His	0	0	3	32
MFSD8 [a]	NM_152778.2:c.670A>T	p.Asn224Tyr	0	0	3	23.9
GUCA1A	NM_000409.4:c.333G>C	p.Glu111Asp	0	0	4	22.9
PRPH2	NM_000322.4:c.611_626del	p.Tyr204Serfs*47	0	0	5	33
PRPH2	NM_000322.4:c.605G>A	p.Gly202Glu	0	0	4	29.8
PRPH2	NM_000322.4:c.512T>G	p.Phe171Cys	0	0	4	27.3
RP1L1	NM_178857.5:c.1024_1026delinsCTCCT	p.Arg342LeufsTer22	0	0	4	22
RP1L1	NM_178857.5:c.196G>C	p.Asp66His	0	0	3	26.6
KCNV2 [c]	NM_133497.3:c.-759_*57289del	p.?	0	0	5	
KCNV2 [a]	NM_133497.3:c.1096del	p.Val366Trpfs*88	0	0	4	13.41
RGR	NM_002921.3:c.236G>A	p.Arg79His	0.0032	0.0141	3	37
BEST1	NM_004183.3:c.907G>T	p.Asp303Tyr	0	0	5	28.5
CNGB1	NM_001297.4:c.2662G>A	p.Ala888Thr	0.0225	0.1145	3	18.28
CNGB1	NM_001297.4:c.1658C>A	p.Ala553Glu	0.0004	0.0009	3	21.1
GUCY2D	NM_000180.3:c.929C>A	p.Thr310Asn	0	0	3	24.5
RS1	NM_000330.3:c.209G>A	p.Gly70Asp	0	0	4	26.7
RS1	NM_000330.3:c.150G>A	p.Trp50*	0	0	5	35
RS1 [d]	NM_000330.3:c.53-717_78+262del	p.Ala18_Glu26delinsGluProGlyGlnHisSerLysThrLeu	0	0	5	
RPGR	NM_001034853.1:c.2819_2838dup	p.Glu947LysfsTer149	0	0	3	22.4

[a] the variant has been found in 2 unrelated probands.

[b] size of the deletion: 3’999-4’033 bp

(NC_000001.10:g.(94507656_94507700)_(94511700_94511710)del)

[c] size of the deletion: 70’036 bp (NC_000009.11:g.2716981_2787016del)

[d] size of the deletion: 1’005 bp (NC_000023.10:g.18675498_18676502del)

Discussion

We have presented a targeted sequencing approach based on LR PCR products that is flexible, versatile, and cost-effective. Primers for 124 LR PCRs to cover 35 RD-associated loci have been established. The method has been validated with 108 molecularly diagnosed RD cases. All previously identified variants could be verified, corresponding to 100% sensitivity. The target regions to be sequenced may be personalized according to the clinical phenotype and family history, so that costs can be reduced compared to standard genetic testing.

Although the method provides the most complete coverage of target loci (Figure 1 and Figure S1), several intronic regions could not be amplified (20 kb out of 1’815 kb, 0.6%). Therefore, few loci (*CRB1*, *EFEMP1*, *IMPG1*, *CNGB3*, and *TIMP3*) have minor gaps in coverage. Moreover, whilst the method is very sensitive to deletions that lie within the amplicons, it is not able to detect deletions spanning primer-binding regions. In this case, ADO would occur and the assay would not show any sign of the CNV other than lower PCR output and the fact that affected regions displays exclusively homozygous-appearing variants.

It is not possible to eliminate the risk of ADO in any PCR; however, we did not observe any such events. Homozygous-appearing regions in non-consanguineous families should warrant caution. A final limitation of the method is the need for high molecular weight DNA as a template.

Since the sequencing targets are selected based on clinical phenotype and family history, this method is highly dependent on precise clinical assessment. Close collaboration between the molecular and the clinical diagnostics teams is therefore vital.

The method was used as a first-tier assay to sequence candidate loci in 94 RD patients (mostly MD) that did not undergo genetic testing previously and as a second-tier assay to discover “second hits” in 25 RD patients of the missing heritability cohort, which resulted in the identification of likely pathogenic variants in 69 probands (58.0%). Published deep-intronic variants in *ABCA4* contributed to the diagnosis of 4 probands (3.4%).

Variants in *ABCA4*, *PRPH2*, and *BEST1* alone explained retinal disease in 60.7% of the MD cohort, similar to previously published results (57%)[11]. Even more striking are the results for the STGD subcohort, where variants in *ABCA4* and *PRPH2* were found to be the likely cause of disease in 76.8% of probands (Figure 2). For this reason, we envision our method to be particularly applicable to diseases with low genetic heterogeneity (such as STGD), and as part of a tiered genetic testing strategy to reduce the number of more costly assays (such as ES) for more genetically heterogeneous diseases (such as unspecified MD and Leber congenital amaurosis). In our study, use of a tiered strategy composed of LR PCR sequencing of *ABCA4*, *PRPH2*, and *BEST1*, followed by ES for the remaining undiagnosed samples, in the MD cohort would have saved 34% in material costs. A similar system was previously found to be more sensitive and less expensive than standard ES analysis in one of the largest RD cohorts yet reported.[10]

Furthermore, the protocol can be useful in analyzing challenging regions that are typically not covered by other methods, such as *RPGR*’s ORF15. It has been shown

previously that NGS of a LR PCR over this region provides good coverage and we have developed a secondary data analysis pipeline to improve sensitivity and specificity.[27,29] This strategy permitted the identification of a novel 20 bp insertion (NM_001034853.1:c.2819_2838dup, Table 2, and Table S4) that was not detected by standard analysis pipelines.

As demonstrated by the examples of the *ABCA4*, *RP1*, *RS1*, and *KCNV2* deletions, the protocol can be adapted easily to characterize the breakpoints of identified CNVs. Even though unambiguous breakpoint characterization was not possible for the *ABCA4* deletion, sequencing revealed the deletion to be flanked by two Alu element and a 10 bp microhomology sequence. The two flanking Alu elements are characterized by variants that increase their homologies, which is suggestive of a gene conversion event.[30] Since the likelihood of these events is indirectly correlated with the distance of the elements, it may be secondary to the deletion event.[30]

As a proof of concept, we showed that finalized LR libraries can be spiked into an exome library to either enhance coverage of poorly captured exonic regions (such as *RPGR*'s ORF15) or to obtain coverage of otherwise uncaptured regions (such as *ABCA4*'s introns) (Figure S5).

Finally, the method can facilitate chromosomal phasing. It allows for segregation analysis of multiple variants on the same locus in a single experiment, when samples from family members are available. Moreover, when family members are not available for testing, read phasing based on the LR PCR sequencing data of the index patient might confirm or exclude compound heterozygosity. However, this depends on the density of heterozygous informative variants in the region of interest (Figure S6).

The method may also be beneficial for other Mendelian diseases, such as *CFTR*-related diseases, Tay-Sachs disease, Marfan syndrome, and tuberous sclerosis.

Data Availability

The following supplementary materials are available online; Figure S1: LR PCRs panel coverage results, Figure S2: *ABCA4* deletion characterization, Figure S3: *KCNV2* deletion characterization, Figure S4: *RS1* deletion characterization, Figure S5: ES spike-in proof of concept, Figure S6: Read phasing, Text S1: Detailed long-range PCR sequencing protocol, Table S1: Long-range PCR primers, Table S2: Long-range PCR primers for breakpoint characterization, Table S3: Sanger sequencing primers, Table S4: Validation cohort findings, Table S5: Molecularly diagnosed No previous testing cohort findings, Table S6: Molecularly diagnosed Missing heritability cohort findings, Table S7: Rare variants found in undiagnosed cases.

Acknowledgments

We are grateful to all probands and family members who participated in this study. Moreover, we would like to thank the Velux Stiftung (project #1371) and all referring physicians. The authors thank Dres. Amit Tiwari, Angela Bahr, and Istvan Magyar for their support. Finally, we appreciate the help provided by Cergentis B.V..

Author Information

Conceptualization, J.M., S.K., W.B.; methodology, J.M.; software, J.M., S.K.; validation, J.M., L.B., and A.M.; formal analysis, J.M.; investigation, J.M., L.B., S.F., F.K.P., J.H., and C.G.K.; resources, W.B. and C.G.K.; data curation, J.M.; writing –

original draft, J.M.; writing – review and editing, J.M., S.K., F.K.P., J.H., C.G.K., and W.B.; visualization, J.M.; supervision, S.K., and W.B.; project administration, J.M. and W.B.; funding acquisition, J.M. and W.B.. All authors have read and agreed to the published version of the manuscript.

Ethics Declaration

The study was conducted in accordance with the Helsinki Declaration. The approval for genetic testing in human patients was awarded to The Institute of Medical Molecular Genetics by the Federal Office of Public Health (FOPH) in Switzerland and written informed consent was obtained from all patients and family members included in this study.

Disclosure

The authors declare no conflict of interest. The sponsor or funding organization had no role in the design or conduct of this research.

References

1. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29(5):335-375. doi:10.1016/j.preteyeres.2010.03.004
2. Sahel JA, Marazova K, Audo I. Clinical characteristics and current therapies for inherited retinal degenerations. *Cold Spring Harb Perspect Med*. 2015;5(2). doi:10.1101/cshperspect.a017111
3. Tiwari A, Lemke J, Altmueller J, et al. Identification of novel and recurrent disease-causing mutations in retinal dystrophies using whole exome

sequencing (WES): Benefits and limitations. *PLoS One*. 2016;11(7). doi:10.1371/journal.pone.0158692

4. Toulis V, Cortés-González V, de Castro-Miró M, et al. Increasing the Genetic Diagnosis Yield in Inherited Retinal Dystrophies: Assigning Pathogenicity to Novel Non-canonical Splice Site Variants. *Genes (Basel)*. 2020;11(4):378. doi:10.3390/genes11040378
5. Weisschuh N, Mayer AK, Strom TM, et al. Mutation detection in patients with retinal dystrophies using targeted next generation sequencing. *PLoS One*. 2016;11(1). doi:10.1371/journal.pone.0145951
6. Huang XF, Huang F, Wu KC, et al. Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. *Genet Med*. 2015;17(4):271-278. doi:10.1038/gim.2014.138
7. Bravo-Gil N, Méndez-Vidal C, Romero-Pérez L, et al. Improving the management of Inherited Retinal Dystrophies by targeted sequencing of a population-specific gene panel. *Sci Rep*. 2016;6(1):1-10. doi:10.1038/srep23910
8. Ellingford JM, Barton S, Bhaskar S, et al. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet*. 2016;53(11):761-767. doi:10.1136/jmedgenet-2016-103837
9. Beryozkin A, Shevah E, Kimchi A, et al. Whole Exome Sequencing Reveals Mutations in Known Retinal Disease Genes in 33 out of 68 Israeli Families with Inherited Retinopathies. *Sci Rep*. 2015;5. doi:10.1038/srep13187
10. Stone EM, Andorf JL, Whitmore SS, et al. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease.

Ophthalmology. 2017;124(9):1314-1331. doi:10.1016/j.ophtha.2017.04.008

11. Birtel J, Eisenberger T, Gliem M, et al. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-018-22096-0
12. Maroilley T, Tarailo-Graovac M. Uncovering Missing Heritability in Rare Diseases. *Genes (Basel)*. 2019;10(4). doi:10.3390/genes10040275
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(25):5136-5145. doi:10.1093/hmg/ddt367
14. Van Cauwenbergh C, Van Schil K, Cannoodt R, et al. ArrEYE: A customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet Med*. 2017;19(4):457-466. doi:10.1038/gim.2016.119
15. 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(8):1761-1771. doi:10.1038/s41436-018-0420-y
16. Ellingford JM, Barton S, Bhaskar S, et al. Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*. 2016;123(5):1143-1150. doi:10.1016/j.ophtha.2016.01.009
17. Zernant J, Xie YA ngel., Ayuso C, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet*. 2014;23(25):6797-6806. doi:10.1093/hmg/ddu396
18. Zernant J, Schubert C, Im KM, et al. Analysis of the ABCA4 gene by next-

generation sequencing. *Investig Ophthalmol Vis Sci*. 2011;52(11):8479-8487. doi:10.1167/iovs.11-8182

19. 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(8):1751-1760. doi:10.1038/s41436-018-0414-9

20. Zernant J, Lee W, Nagasaki T, et al. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol case Stud*. 2018;4(4). doi:10.1101/mcs.a002733

21. 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(3):236-246. doi:10.1038/ng0397-236

22. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13:134. doi:10.1186/1471-2105-13-134

23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

24. Depristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-501. doi:10.1038/ng.806

25. Povysil G, Tzika A, Vogt J, et al. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum Mutat*. 2017;38(7):889-897. doi:10.1002/humu.23237

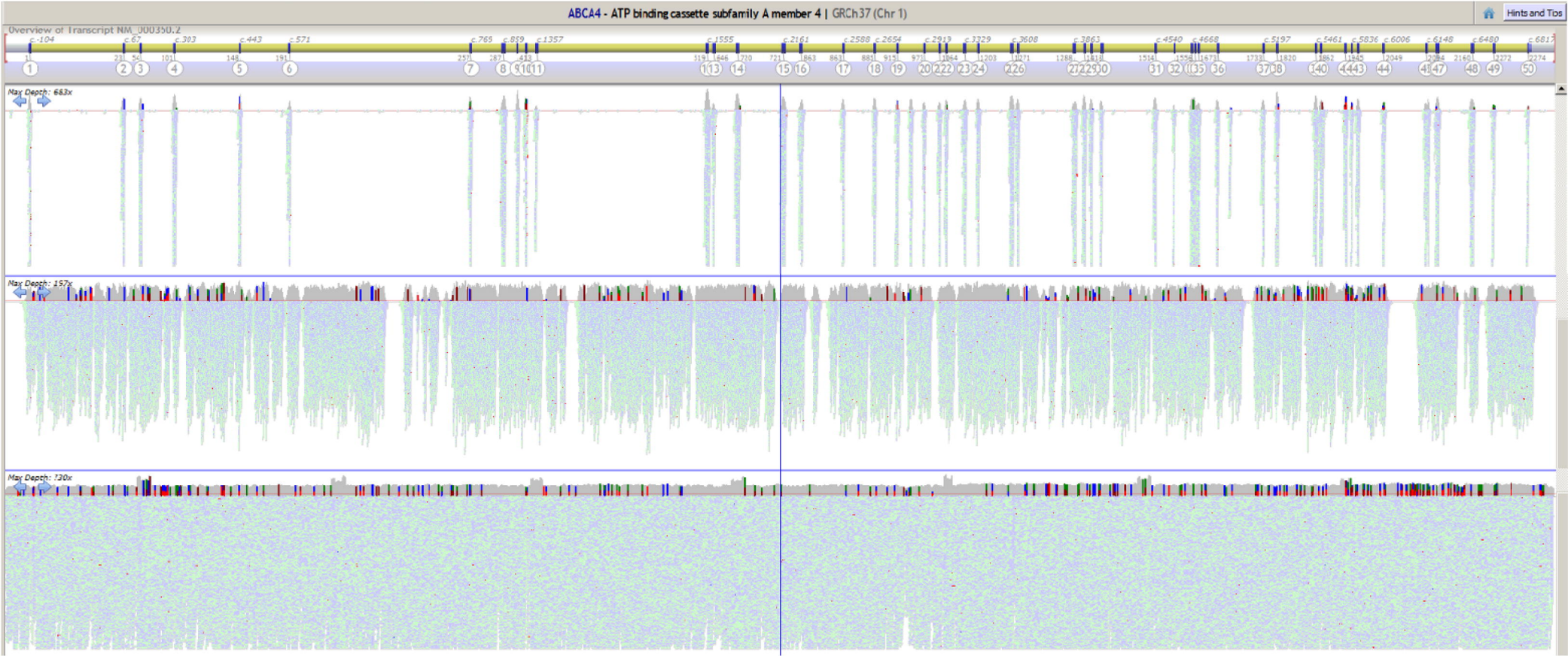
26. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021

27. Maggi J, Roberts L, Koller S, Rebello G, Berger W, Ramesar R. De Novo Assembly-Based Analysis of RPGR Exon ORF15 in an Indigenous African Cohort Overcomes Limitations of a Standard Next-Generation Sequencing (NGS) Data Analysis Pipeline. *Genes (Basel)*. 2020;11(7):35-50. doi:10.3390/genes11070800

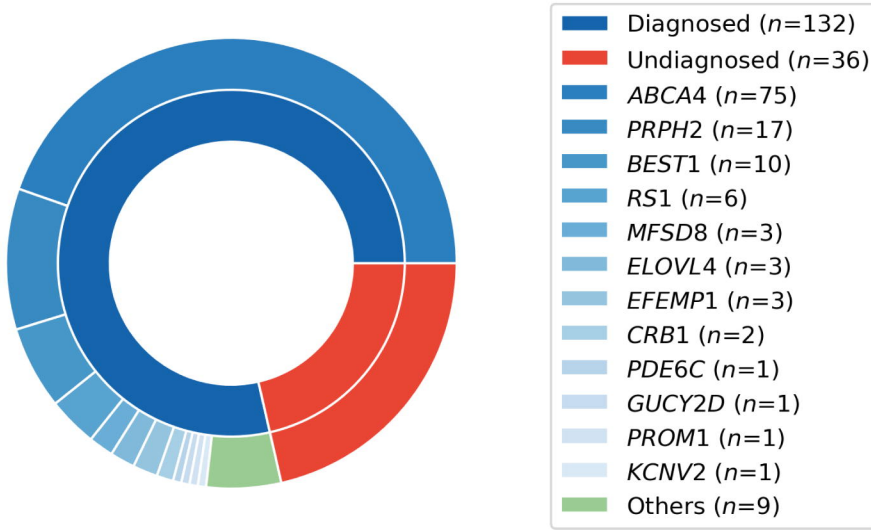
28. Carss K, Arno G, Erwood M, et al. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet*. 2017;100(1):75-90. doi:10.1016/j.ajhg.2016.12.003

29. Li J, Tang J, Feng Y, et al. Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J Mol Diagnostics*. 2016;18(6):817-824. doi:10.1016/j.jmoldx.2016.06.007

30. Zhi D. Sequence correlation between neighboring Alu instances suggests post-retrotransposition sequence exchange due to Alu gene conversion. *Gene*. 2007;390(1-2):117-121. doi:10.1016/j.gene.2006.09.030


ABCA4 - ATP binding cassette subfamily A member 4 | GRCh37 (Chr 1)
Hints and Tips
Overview of Transcript NM_000350.2
Max Depth: 683x
Max Depth: 157x
Max Depth: 730x

Macular dystrophies



Stargardt disease

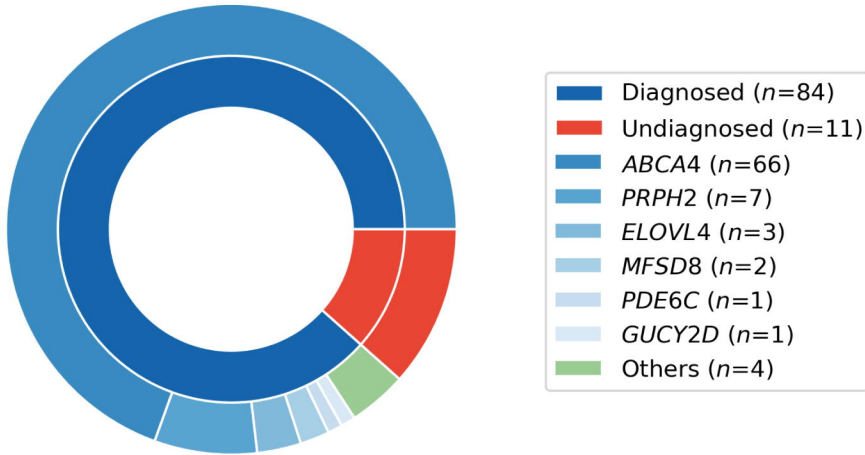


Table 1. **Established long-range PCRs for retinal diseases-associated loci**. The first and second columns report the corresponding transcript identifiers, respectively, of the loci included in the panel, ordered by chromosome coordinates. The third column highlights the number of PCRs necessary to cover the locus comprehensively, and the last column shows the total size of genomic target sequence.

Locus	Transcript	# of PCRs needed	Genomic target size (bp)
PPT1	NM_000310.3	2	26’746
ABCA4	NM_000350.2	8	137’748
CRB1	NM_201253.2	14	212’898
PCARE	NM_001029883.2	1	18’628
EFEMP1	NM_001039348.2	5	60’569
IMPG2	NM_016247.3	6	94’768
PROM1	NM_006017.2	7	117’712
MFSD8	NM_152778.2	3	49’747
CTNNA1	NM_001323982.1	13	182’694
GUCA1A	NM_000409.4	2	27’191
GUCA1B	NM_002098.5	1	13’548
PRPH2	NM_000322.4	2	32’242
IMPG1	NM_001563.3	11	144’700
ELOVL4	NM_022726.3	2	34’792
DHS6S1	NA	1	13’473
RP1L1	NM_178857.5	3	51’877
RP1	NM_006269.1	1	16’930
CNGB3	NM_019098.4	12	172’052
KCNV2	NM_133497.3	1	15’214
ATOH7	NM_145178.3	1	9’275
PDE6C	NM_006204.3	3	53’794
BEST1	NM_004183.3	1	17’272
C1QTNF5	NM_015645.4	1	10’536
PDE6H	NM_006205.2	1	11’084
RDH5	NM_002905.3	1	6’122
OTX2	NM_021728.3	1	12’758
NR2E3	NM_014249.3	1	11’097
RLBP1	NM_000326.4	1	13’667
GUCY2D	NM_000180.3	1	18’939
FSCN2	NM_001077182.2	1	11’574
RAX2	NM_001319074.1	1	5’845
TIMP3	NM_000362.4	5	67’146
RS1	NM_000330.3	3	34’731
RPGR	NM_001034853.1	4	60’808
RP2	NM_006915.2	3	46’521

Table 1. Macular dystrophies cohort summary

Category	Tot. #	Tot. diagnosed	Tot .undiagnosed	Mean age at referral (years)	Diagnostic yield (%)
Overall	168	132	36	37.86	78.6
Stratification					
Validation	70	70	0	35.60	100.0
Missing heritability	13	4	9	35.08	30.8
No previous testing	85	58	27	40.14	68.2
Clinical diagnosis					
STGD	95	84	11	35.88	88.4
Unspecified MD	45	28	17	40.11	62.2
BEST	11	9	2	35.73	81.8
ML	9	5	4	47.89	55.6
XLRS	6	6	0	32.00	100.0
SFD	1	0	1	63.00	0.0
AVMD	1	0	1	67.00	0.0

Table 3. Novel likely pathogenic variants

Locus	cNomen	pNomen	gnomAD overall (%)	gnomAD max. (%)	ACMG class	CADD score
ABCA4	NM_000350.2:c.6731T>A	p.Val2244Glu	0	0	3	27.3
ABCA4	NM_000350.2:c.6428T>A	p.Met2143Lys	0	0	4	32.0
ABCA4	NM_000350.2:c.6323_6331delinsGGC	p.Met2108_Asn2111delinsArgHis	0	0	4	35
ABCA4	NM_000350.2:c.5924G>T	p.Gly1975Val	0	0	4	28.8
ABCA4	NM_000350.2:c.5690_5704del	p.Gln1897_Phe1901del	0	0	3	22.2
ABCA4	NM_000350.2:c.5691G>T	p.Gln1897His	0	0	3	23.8
ABCA4	NM_000350.2:c.5461-6T>C	p.?	0	0	3	14.93
ABCA4 [a]	NM_000350.2:c.4958G>A	p.Gly1653Glu	0	0	3	28.1
ABCA4	NM_000350.2:c.4609del	p.Thr1537Argfs*6	0	0	4	33
ABCA4	NM_000350.2:c.4383G>C	p.Trp1461Cys	0.0004	0.003	3	32
ABCA4	NM_000350.2:c.3323del	p.Arg1108Profs*40	0	0	5	34
ABCA4	NM_000350.2:c.3179A>C	p.Gln1060Pro	0.0008	0.0065	4	23.7
ABCA4 [b]	NM_000350.2:c.(2918+765_2918+775)_(3328+618_3328+662)del	p.Leu973_Asp2273delinsPheMetAlaArgValGluArgSerLeuGlyAsn	0	0	5	
ABCA4	NM_000350.2:c.1742C>A	p.Thr581Asn	0	0	3	26.7
ABCA4	NM_000350.2:c.1621_1622del	p.Leu541Thrfs*14	0	0	4	32
ABCA4	NM_000350.2:c.727_728dup	p.Tyr245Cysfs*18	0	0	5	26.2
ABCA4	NM_000350.2:c.676C>A	p.Arg226Ser	0.0068	0.0163	3	14.13
CRB1	NM_201253.2:c.1472A>T	p.Asp491Val	0	0	3	15.86
CRB1	NM_201253.2:c.2298G>A	p.Trp766*	0	0	5	36
OPA1	NM_130837.2:c.2987A>C	p.Lys996Thr	0	0	4	23.8
PROM1	NM_006017.2:c.2476G>C	p.Asp826His	0	0	3	32
MFSD8 [a]	NM_152778.2:c.670A>T	p.Asn224Tyr	0	0	3	23.9
GUCA1A	NM_000409.4:c.333G>C	p.Glu111Asp	0	0	4	22.9
PRPH2	NM_000322.4:c.611_626del	p.Tyr204Serfs*47	0	0	5	33
PRPH2	NM_000322.4:c.605G>A	p.Gly202Glu	0	0	4	29.8
PRPH2	NM_000322.4:c.512T>G	p.Phe171Cys	0	0	4	27.3
RP1L1	NM_178857.5:c.1024_1026delinsCTCCT	p.Arg342LeufsTer22	0	0	4	22
RP1L1	NM_178857.5:c.196G>C	p.Asp66His	0	0	3	26.6
KCNV2 [c]	NM_133497.3:c.-759_*57289del	p.?	0	0	5	
KCNV2 [a]	NM_133497.3:c.1096del	p.Val366Trpfs*88	0	0	4	13.41
RGR	NM_002921.3:c.236G>A	p.Arg79His	0.0032	0.0141	3	37
BEST1	NM_004183.3:c.907G>T	p.Asp303Tyr	0	0	5	28.5
CNGB1	NM_001297.4:c.2662G>A	p.Ala888Thr	0.0225	0.1145	3	18.28
CNGB1	NM_001297.4:c.1658C>A	p.Ala553Glu	0.0004	0.0009	3	21.1
GUCY2D	NM_000180.3:c.929C>A	p.Thr310Asn	0	0	3	24.5
RS1	NM_000330.3:c.209G>A	p.Gly70Asp	0	0	4	26.7
RS1	NM_000330.3:c.150G>A	p.Trp50*	0	0	5	35
RS1 [d]	NM_000330.3:c.53-717_78+262del	p.Ala18_Glu26delinsGluProGlyGlnHisSerLysThrLeu	0	0	5	
RPGR	NM_001034853.1:c.2819_2838dup	p.Glu947LysfsTer149	0	0	3	22.4

[a] the variant has been found in 2 unrelated probands.

[b] size of the deletion: 3’999-4’033 bp (NC_000001.10:g.(94507656_94507700)_(94511700_94511710)del)

[c] size of the deletion: 70’036 bp (NC_000009.11:g.2716981_2787016del)

[d] size of the deletion: 1’005 bp (NC_000023.10:g.18675498_18676502del)