

Massively parallel sequencing for early molecular diagnosis in Leber congenital amaurosis

Frauke Coppieters, MSc, PhD¹, Bram De Wilde, MSc, MD¹, Steve Lefever, MSc¹, Ellen De Meester, MSc², Nina De Rocker, MSc¹, Caroline Van Cauwenbergh, MSc¹, Filip Pattyn, MSc, PhD¹, Françoise Meire, MD, PhD³, Bart P. Leroy, MD, PhD^{1,4}, Jan Hellemans, MSc, PhD¹, Jo Vandesompele, MSc, PhD¹ and Elfride De Baere, MD, PhD¹

Purpose: Leber congenital amaurosis (LCA) is a rare congenital retinal dystrophy associated with 16 genes. Recent breakthroughs in LCA gene therapy offer the first prospect of treating inherited blindness, which requires an unequivocal and early molecular diagnosis. While present genetic tests do not address this due to a tremendous genetic heterogeneity, massively parallel sequencing (MPS) strategies might bring a solution. Here, we developed a comprehensive molecular test for LCA based on targeted MPS of all exons of 16 known LCA genes.

Methods: We designed a unique and flexible workflow for targeted resequencing of all 236 exons from 16 LCA genes based on quantitative PCR (qPCR) amplicon ligation, shearing, and parallel sequencing of multiple patients on a single lane of a short-read

sequencer. Twenty-two prescreened LCA patients were included, five of whom had a known molecular cause.

Results: Validation of 107 variations was performed as proof of concept. In addition, the causal genetic defect and a single heterozygous mutation were identified in 3 and 5, respectively, of 17 patients without previously identified mutations.

Conclusion: We propose a novel targeted MPS-based approach that is suitable for accurate, fast, and cost-effective early molecular testing in LCA, and easily applicable in other genetic disorders.

Genet Med 2012;14(6):576–585

Key Words: Leber congenital amaurosis; massively parallel sequencing; molecular diagnosis; qPCR

INTRODUCTION

Leber congenital amaurosis (LCA) (OMIM no. 204000) represents the earliest and most severe autosomal recessive retinal dystrophy, causing profound visual deficiency or blindness from birth. LCA has a worldwide incidence of ~1/30,000 and is the most frequent cause of childhood blindness. Patients are diagnosed in their first year of life based on the presence of severe visual loss, a nondetectable or strongly impaired electroretinogram, and nystagmus.^{1,2}

As for all retinal dystrophies, the progressive retinal degeneration leading to blindness has been irreversible thus far. However, in 2008, three independent phase I clinical trials achieved a major breakthrough following gene-replacement therapy in three young adults with *RPE65*-related LCA, which was shown to be safe and resulted in visual improvement in a subset of patients.^{3–5} A follow-up study including children demonstrated even more beneficial effects at a younger age.⁶ The efficacy and safety persisted through up to 2 years.^{6–8}

Despite this tremendous step forward, a major obstacle remains in identifying LCA patients eligible for gene-specific treatment due to a massive genetic heterogeneity. Sixteen disease genes are known to account for ~70% of LCA cases, leaving the remaining 30% unexplained (RetNet, <http://www.sph.uth.tmc.edu/>

RetNet).⁹ Moreover, some of these genes are also involved in syndromic diseases.^{10,11} Consequently, an early molecular diagnosis offers the prospect of specific and adequate medical follow-up.

To date, the most commonly used genetic test for LCA is a microarray evaluating 641 known variants in 13 genes.¹² Although this technique represents a good first-pass screening, it fails in detecting new mutations, is expensive in a routine context, and has a variable, population-dependent detection rate.^{13–15} Secondary genetic tests comprise denaturing high-performance liquid chromatography and Sanger sequencing, mostly of a subset of genes because of excessive costs and workload.^{1,16,17} Hence, an urgent need exists for an effective approach able to identify mutations in all currently known LCA genes.

Massively parallel sequencing (MPS) technologies are highly suitable for molecular diagnosis of genetically heterogeneous disorders, given their high throughput and decreasing sequencing cost.¹⁸ However, a major challenge remains to enrich regions of interest, such as exons, from the patient's genome. Although hybridization-based DNA capture is often used, this approach has several limitations, including a high cost, suboptimal capturing efficiency at repetitive regions, interference from homologous sequences, a large variation in coverage, and lack of flexibility if new regions need to be captured.^{19–21} Only a few groups

¹Center for Medical Genetics Ghent, Ghent University, Ghent, Belgium; ²Department of Analytical Chemistry, Ghent University, Ghent, Belgium; ³Department of Ophthalmology, Hôpital des Enfants Reine Fabiola, Brussels, Belgium; ⁴Department of Ophthalmology, Ghent University Hospital, Ghent, Belgium. Elfride De Baere (Elfride.DeBaere@UGent.be)

Submitted 7 August 2011; accepted 1 November 2011; advance online publication 26 January 2012. doi:10.1038/gim.2011.51

have designed PCR-based enrichment protocols, mostly using long-range PCR.^{22–25} Following amplification, all products are quantified and normalized to equimolar amounts, which is time-consuming when handling large numbers of samples.

The goal of this study was to design an accurate, fast, and cost-efficient tool for molecular testing of all known LCA genes using MPS. For this purpose, we applied a novel quantitative PCR (qPCR)-based enrichment strategy that overcomes the aforementioned issues. We tested this LCA panel in 22 LCA patients, performed a thorough variant validation as proof of concept, and identified causal mutations in a subset of prescreened patients.

MATERIALS AND METHODS

Patients

This study was conducted in accordance with the Declaration of Helsinki. Twenty-two sporadic LCA patients were included. The first part of this study was mainly intended for validation and comprised 10 patients (patients 1–10); the second part of this study was an application of the workflow in 12 additional patients without previously identified mutations (patients 11–22).

For the first part, the causal genetic defect was known for five patients (so-called positive controls; patients 4, 5, 6, 9, and 10), whereas no mutations had yet been identified in the remaining five (patients 1, 2, 3, 7, and 8). All patients previously underwent microarray testing (LCA chip versions 2004–2010; Asper Ophthalmics). In the positive control patients, Sanger sequencing was performed of *CEP290*, *CRB1*, *RPE65*, *GUCY2D*, *AIPL1*, and *CRX*.¹⁶ In addition, patients 4 and 5, originating from a consanguineous marriage, were sequenced for *IQCB1* and *RDH12*, respectively, following identification of these genes in homozygous regions (Affymetrix GeneChip Human Mapping 250K arrays; Affymetrix, Santa Clara, CA).

Prescreening of the 12 patients included in the second part of this study consisted of LCA chip analysis in all patients and sequencing of *CEP290*, *CRB1*, *RPE65*, *GUCY2D*, *AIPL1*, and *CRX* in patients 18–22.¹⁶

More than 160 unrelated healthy individuals were used as a control panel. Genomic DNA was extracted from leukocytes using the Puregene DNA isolation kit (Gentra, Minneapolis, MN).

Enrichment

Primers were designed using the in-house primerXL pipeline to cover all 252 exons of *RD3*, *RPE65*, *CRB1*, *MERTK*, *IQCB1*, *LRAT*, *LCA5*, *TULP1*, *IMPDH1*, *CEP290*, *RPGRIP1*, *RDH12*, *SPATA7*, *AIPL1*, *GUCY2D*, and *CRX* and the deep intronic *CEP290* mutation c.2991+1655A>G (Ensembl, Release 55, GRCh37) (S. Lefever, F. Pattyn, B. De Wilde *et al.*, unpublished data). Two primer designs were performed (maximal amplicon length of 400 and 600 bp). From these designs, a selection was made to obtain an overall coverage of all target bases as efficiently as possible.

qPCRs were prepared in 384-well plates with a 96-well head pipetting robot (Tecan Freedom Evo 100; Tecan, Männedorf,

Switzerland) using the SsoFast EvaGreen Supermix (Bio-Rad, Nazareth Eke, Belgium). Subsequently, amplicons from one patient were pooled, ligated, and sheared. Parameters of the primer design and conditions of the qPCR, ligation, and shearing reactions are specified in **Supplementary Methods and Procedures** online.

Sequencing on the Illumina Genome Analyzer IIx

Library preparations and sequencing were performed by the Ghent University NXTGNT consortium following manufacturer's protocols. All patients were uniquely tagged using the Multiplexing Sample Preparation Oligonucleotide kit (Illumina, Eindhoven, the Netherlands). Of the 12 LCA patients pooled in the first lane, 10 were used for validation of the protocol (lane 1, patients 1–10). Libraries were prepared from a 300-bp size-selected fraction, subjected to 18 cycles of PCR and evaluated for their size (2100 Bioanalyzer, DNA 1000 kit; Agilent, Diegem, Belgium) as well as concentration (Quant-iT PicoGreen dsDNA assay; Invitrogen, Ghent, Belgium). Finally, 120 μ l of a pool of the 12 normalized libraries (concentration of 7 pmolar) was subjected to single-end sequencing of 100 cycles.

For the second lane, libraries were prepared from a 200–300-bp size-selected fraction, and library quantification was performed using qPCR following manufacturer's protocols (lane 2, patients 11–22). In total, 120 μ l of a pool of the 12 normalized libraries (concentration of 10 pmolar) was subjected to paired-end sequencing of 2 \times 45 cycles.

Read mapping and variant analysis

Image analysis and base calling was performed using the Genome Analyzer Pipeline Software (Illumina). Sequence analysis was carried out with the NextGENe software v2.00 (SoftGenetics, LLC, State College, PA). Reads with a median quality score <20 or with ambiguous nucleotide calls ($N \geq 3$) were removed. Subsequently, reads were aligned against GenBank reference sequences (**Supplementary Table S1** online).

Two distinct condensation methods were used: error correction and consolidation (differences specified in **Supplementary Methods and Procedures** online). The latter merges overlapping sequences and uses consensus sequences instead of the original reads. Default alignment settings were applied, with two exceptions: the minimal coverage for a mutation to be called was set to 1 and the “detect structural variations” option was selected. For lane 2, paired reads were taken into account during error correction mapping.

Coverage data were obtained for all exons with 20-bp up- and downstream intronic sequence. Variant nomenclature uses numbering with the A of the initiation codon ATG as +1 (<http://www.hgvs.org/mutnomen>), based on the RefSeqs listed in **Supplementary Table S1** online.

The pathogenic potential of novel variants was assessed using the Alamut mutation interpretation software v1.54 (Interactive Biosoftware, Rouen, France), complemented by the NetGene2 Server²⁶ and Berkeley Drosophila Genome Project splice-site prediction tools.²⁷

RESULTS

We present a high-throughput molecular test for all currently known LCA genes using a novel enrichment strategy based on qPCR, ligation, and fragmentation, followed by sequencing on the Illumina Genome Analyzer IIx. The first part of this study was mainly aimed at validation of the enrichment protocol (proof of concept, patients 1–10), whereas the second part of this study consisted of a blind screening of 12 prescreened mutation-negative patients with LCA (patients 11–22).

Enrichment of LCA disease genes

qPCR was used to target all exons from 16 LCA disease genes. In total, 375 qPCR amplicons were designed, which together amplify ~152 kb. As a validation, qPCR parameters were evaluated in patients 1–10 (Figure 1 and Supplementary Figure S1 online). More than 80% of amplicons generated a quantification cycle (C_q) value between 23 and 27 (Figure 1a). In addition, end point fluorescence values ranged from 40 to 65 in more than 86% of amplicons, pointing at very similar end concentrations (Figure 1b).

Cycling conditions were adjusted to allow amplification of fragments with a length ranging from 118 to 783 bp. No effect of amplicon size was observed (Supplementary Figure S1a,b). However, qPCR amplification was influenced by the amplicon GC content and secondary structure Gibbs free energy (*dG*) (Supplementary Figure S1c–f).

Supplementary Figure S2 shows ligation and shearing results. As only a fraction of unligated PCR product was present in comparison with the amount of ligated DNA, no additional purification was performed (Figure 1c).

Sequencing output and coverage

Supplementary Tables S2 and S3 online summarize the results of quality filtering, condensation, and alignment steps

performed by NextGENe for lanes 1 and 2. The wide variability in the number of reads between patients included in lane 1 can be attributed to the picogreen-based library quantification method, which was replaced by qPCR quantification for lane 2.

Further coverage and variant analysis was based on the National Center for Biotechnology Information RefSeq sequences and the 236 (224 coding) exons defined herein (Supplementary Table S1). Figure 2 shows cumulative percentage plots of the average and minimal coverage for the original as well as condensed reads for all exons, including 20-bp up- and downstream intronic sequences.

The expected average coverage of lane 1 was calculated to be 918-fold per patient, in case of 200,000 clusters/tile with a 70% pass filter rate. Due to a lower number of clusters and wide distribution of reads between patients (Figure 2), however, this expected coverage was achieved only for patients 1, 2, 7, and 8. For these patients, more than 90% of exons were covered 40 times or more (Supplementary Table S2).

Improved specifications for lane 2 resulted in an expected average coverage of 1,239 (300,000 clusters/tile with a 70% pass filter rate), which was exceeded in all patients. More than 90% of exons had a minimal coverage of at least 40 in patients 11 and 16–22 (Supplementary Table S3).

Overall, 17 exons lacked sufficient coverage depth in all patients. These exons correspond with 23 distinct amplicons, of which 18 amplicons (5%) resulted in zero coverage for the whole or part of the exon, presumably due to high GC content and/or low *dG* (Supplementary Table S4).

Variant analysis and interpretation

Validation of the protocol using previously identified variants. Validation of our protocol consisted of the assessment of all sequence variants, both mutations and polymorphisms, that were previously identified in patients 1–10.

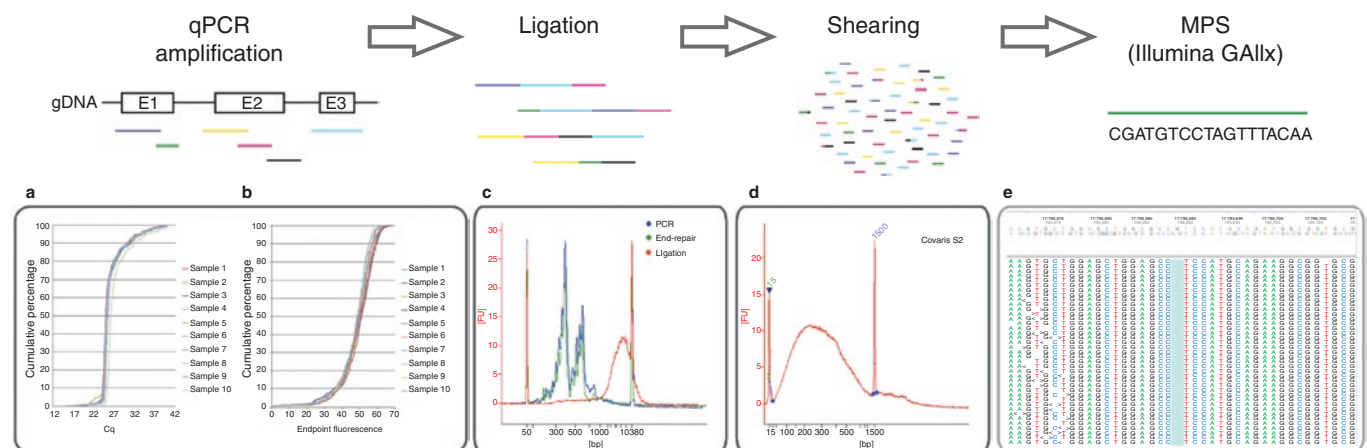


Figure 1 Workflow overview. For each patient, all exons and intron–exon boundaries of the 16 known LCA genes are amplified using qPCR, followed by random ligation and shearing. Subsequently, 12 amplicon pools are indexed and sequenced in one lane of the Illumina Genome Analyzer IIx. (a) Cumulative distribution plot of mean C_q values of both replicates for each patient of lane 1. (b) Cumulative distribution plot of mean end-point fluorescence values of both replicates for each patient of lane 1. (c) Merged Agilent Bioanalyzer electropherograms (DNA 7500) of purified pooled PCR-product (blue, 1/4 dilution), purified end-repaired DNA (green, 1/4 dilution), and ligated DNA (red, 1/3 dilution) from patient 7. (d) Agilent Bioanalyzer electropherogram (DNA 1000) of sheared product from patient 7. (e) Screenshot of the NextGENe alignment viewer showing the homozygous *AIP1* mutation c.885del, identified in patient 17. C_q, quantification cycle; LCA, Leber congenital amaurosis; qPCR, quantitative PCR.

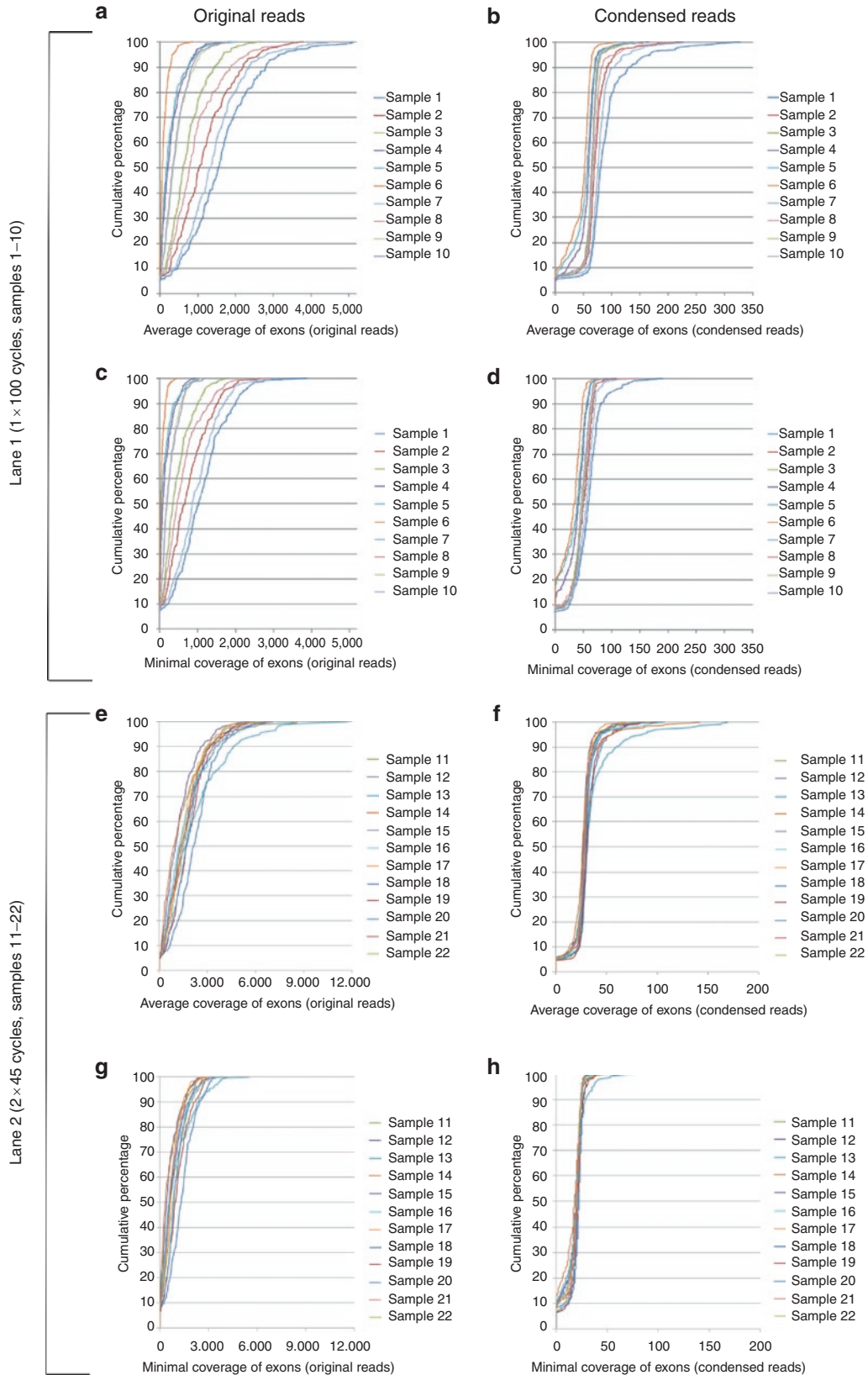


Figure 2 Distribution of average and minimal coverage for both original and condensed reads. Coverage data were obtained using .bed files containing all exons with 20-bp up- and downstream intronic sequence.

Table 1 Evaluation of seven mutations previously identified by Sanger sequencing, using consolidation as well as error-correction methods

Patient	Gene	Human Genome Variation Society mutation nomenclature	NextGENE mutation call	Hom/Het	Mapping with consolidation			Mapping with error correction		
					Cov	Score	Freq	Cov	Score	Freq
4	<i>IQCB1</i>	c.1074_1075dup (p.Ala359GlufsX3)	c.1075_1076insAG (FS)	Hom	59	30	93.2	104	30	89.4
5	<i>RDH12</i>	c.912G>A (Trp304X)	c.[912G>A]+[912G>A] (304W>X)	Hom	63	30	100	249	30	99.6
6	<i>CRB1</i>	c.2441_2442del (p.Leu814ArgfsX23)	c.2441_2442delTA (FS)	Het	65	30	26.2	98	29	39.8
		c.3713_3716dup (p.Cys1240ProfsX24)	c.3716_3717insGCCT (FS)	Het	61	30	31.2	132	28	34.9
9	<i>CEP290</i>	c.2991+1655A>G (p.Cys998X)	c.[2991+1655A>G]+[=]	Het	31	30	45.2	390	30	44.4
		c.5865_5867delAGAinsGG (p.Glu1956GlyfsX9)	c.5867delA (FS)	Het	34	30	52.9	615	30	41.6
10	<i>RPE65</i>	c.991_993dup (p.Trp331dup)	c.993_994insTGG (FS)	Hom	60	30	96.7	344	30	54.1

Cov, coverage; Freq, frequency; FS, frameshift; Het, heterozygous; Hom, homozygous.

First, the presence of seven mutations was evaluated (patients 4–6, 9, and 10). All of these known mutations were detected by NextGENE (Table 1). The frequency of the homozygous mutations varied between 93% and 100%. The frequency of the heterozygous changes approximated 50% for patient 9 but was quite low for both mutations present in patient 6, despite sufficient coverage (Table 1). Strikingly, the frequency of both these heterozygous variants was higher when considering mapping with error correction as the condensation method (Table 1). For c.3713_3716dup, this could partially be explained by the presence of reads starting or ending at the duplicated GCCT site in the consolidation mapping project (Supplementary Figure S3 online). Importantly, the frequency of the other mutations was not increased in the error-corrected mapping project in comparison with the consolidation mapping project (Table 1).

Second, the protocol was validated by the evaluation of 100 (39 distinct) polymorphisms previously identified, including 42 heterozygous and 58 homozygous variants located within the exons and 200-bp up- and downstream sequence (Supplementary Table S5). All variants were identified in the consolidation project, except for three representing two distinct polymorphisms. The first one is c.907-16_907-14del (*RPGRIP1*), which was present in patients 2 (heterozygous), 5 (homozygous), 6 (heterozygous), and 10 (heterozygous) according to the LCA chip. However, NextGENE identified this variant only in patients 2 and 10 (Supplementary Table S5). For patient 5, the variant was present in one read, but not included in the mutation report. Again, when considering the original data in the error-corrected mapping project, the c.907-16_907-14del variant was visible in both patients, albeit in only one read in patient 6 (Supplementary Figure S4a). The frequency of 26% of this variant in the error-correction project of patient 5 questioned the homozygous call of the LCA chip. Indeed, Sanger

sequencing confirmed a heterozygous c.907-16_907-14del variant. Although the repetitive nature of the region might hamper correct detection, an important discrepancy for the variant frequency remains for patient 5 between the two condensation methods. A similar observation was made for c.2818-50G>C (*CEP290*), which was previously identified by Sanger sequencing in a heterozygous state in patient 5. Although this variant was present in 20% of reads in the error-corrected mapping project, it was not reported in the consolidation mapping project (Supplementary Figure S4b).

In addition to c.907-16_907-14del (*RPGRIP1*), two polymorphisms displayed a frequency <25% (Figure 3). They were both detected in a heterozygous state in patient 5, having the second lowest number of reads, and were located in a homopolymeric tract of 7 (c.4704+46del) and 8 (c.3574-9del) thymines in *CEP290* (Supplementary Table S5).

The mutation nomenclature handled by NextGENE approximated Human Genome Variation Society guidelines (<http://www.hgvs.org>), as shown in Tables 1 and 2. One important inconsistency was noted. For the c.2441_2442del mutation, NextGENE reported two variants: one at position 2441 (c.2441_2442delTA) and one at position 2442 (no mutation call). At position 2441, a single-nucleotide polymorphism has been described, corresponding to a missense variant (rs62636268). NextGENE included this polymorphism in the information for the c.2441_2442delTA call. Strikingly, the c.2441_2442delTA was not listed anymore when reported variants were excluded from the mutation report. However, NextGENE still included a c.2442delA variant, occurring at the second position of the deletion.

Identification of mutations in mutation-negative patients. This study included 17 patients in which previous Sanger sequencing of six genes and/or LCA chip analysis did not identify causal mutations (lane 1: patients 1–3, 7, and 8; lane 2: patients 11–22). Given the exclusion of 3 of 107 known variants by the

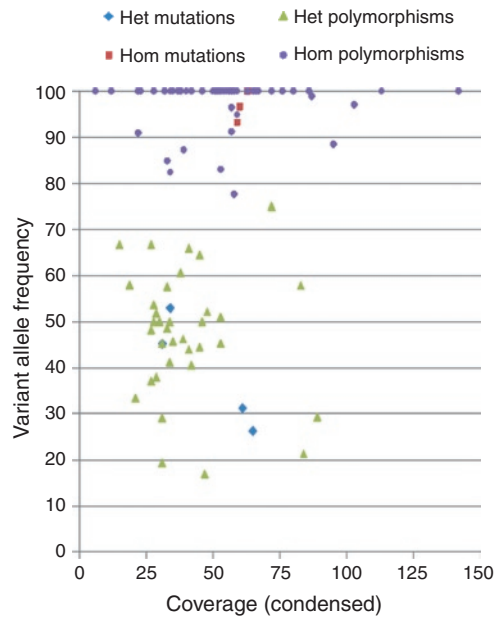


Figure 3 Distribution of variant allele frequencies of 7 mutations and 97 polymorphisms previously identified. Taking into account all 104 variants reported by NextGENe, the mean frequencies for heterozygous and homozygous changes were 46.4 (s.d. = 12.9) and 97.6 (s.d. = 5.2), respectively. Het, heterozygous; Hom, homozygous.

consolidation method of condensation, variant analysis was performed using error correction, in which the original reads are maintained. To search for potential mutations in these patients, variants were first selected based on their coverage and variant allele frequency. Subsequently, variants were evaluated for their potential pathogenic effect using Alamut, their frequency in dbSNP (version 132) and the other patients, and their presence in literature and/or locus-specific mutation databases (<http://www.retina-international.org/sci-news/mutation.htm>). All potential mutations were confirmed through Sanger sequencing of the involved amplicon (**Supplementary Table S6**).

This approach allowed the identification of the causal genetic defect in three patients (**Table 2**). First, patient 7 was found to be heterozygous for c.2302C>T (p.Arg768Trp) and c.2182G>A (p.Asp728Asn) in *GUCY2D*. Both variants affect a strongly conserved nucleotide and amino acid, are predicted to affect protein function, and are described as disease-causing variations (<https://www.carverlab.org/database>). Moreover, p.Arg768Trp is presumably a founder mutation in the northwest of Europe.¹⁵ This mutation was actually a missed call on the LCA chip. Second, a novel homozygous missense mutation in *RDH12*, c.176T>G (p.Leu59Arg), was identified in patient 13. This mutation affects a highly conserved nucleotide and amino acid, is predicted to affect protein function, and is located in the dehydrogenase/reductase domain. Third, a novel homozygous frameshift mutation in *AIPL1*, c.885del (p.Ser296ProfsX10), was found in patient 17. This patient is also a heterozygous carrier of c.472G>A (p.Ala158Thr) in *CRX*, which has previously been described in cone-rod dystrophy and LCA, as well as in 2/110 control individuals.^{28,29} The presence of the A allele

in 12.7% of the Yoruba population (1000Genomes Project, pilot_1_YRI_low_coverage_panel) points to a polymorphism rather than a mutation (rs61748445). In addition, patient 2 carries two *AIPL1* missense variants of which the pathogenic effect is currently uncertain: c.140C>G (p.Thr47Arg) and c.937G>T (p.Ala313Ser) (**Table 2**). Both variants were absent in >160 control individuals.

Moreover, six patients were heterozygous carriers of a single potential pathogenic variant, without a second mutation identified following Sanger sequencing of all exons lacking sufficient coverage. First, a heterozygous c.2577G>T (*P.=*) variant in *GUCY2D* was identified in patient 1. This substitution affects the first nucleotide of exon 14 and is predicted to cause loss of the wild-type acceptor splice site by Alamut, NetGene2,²⁶ and the Berkeley Drosophila Genome Project.²⁷ The T allele was found in 1.7% of the Yoruba population (1000Genomes project, pilot_1_YRI_low_coverage_panel) and in 0.6% of 4,548 chromosomes of the National Heart, Lung, and Blood Institute's Exome Sequencing Project (rs112372281). In addition, this variant was absent in >160 control individuals (own data). Interestingly, patient 13 is also heterozygous for this variant. Second, patient 14 is heterozygous for the novel c.1441G>A (p.Glu481Lys) mutation in *IQCB1*, which affects a highly and moderately conserved nucleotide and amino acid, respectively, and is predicted to affect protein function. Third, two novel heterozygous missense variants in *MERTK* were identified in patients 11 (c.1893C>G, p.Ile631Met) and 21 (c.2237A>G, p.Lys746Arg), respectively. Both variants affect a highly conserved nucleotide and amino acid, are predicted to affect protein function, and are located in the tyrosine kinase domain. Finally, a single novel heterozygous unclassified variant was found in patient 18 (c.3773C>T, p.Thr1258Ile; *RPGRIP1*) and patient 20 (c.461_463del, p.Glu154del; *CEP290*) (**Table 2**).

DISCUSSION

The goal of this study was to design and validate a comprehensive, accurate, and affordable molecular test for LCA. To this end, we developed an innovative workflow for fast and flexible enrichment of a large number of genes before MPS. Our straightforward approach is based on high-throughput qPCR amplification, ligation, and shearing, thus enabling sequencing of regions of interest with variable length, such as exons, on a short-read sequencer (**Figure 1**).

Our workflow holds a number of technical advantages as compared with current molecular tests for retinal dystrophies based on large-scale resequencing. So far, five groups have developed a molecular test for retinal disease genes using custom-made resequencing chips^{30,31} or MPS.^{32–34} Three of these either employ PCR-based enrichment followed by product quantification^{30,34} or use a PCR efficiency MPS run³² to obtain equimolar products prior to sequencing. However, the combination of our in-house primerXL pipeline and highly efficient qPCR amplification eliminates the need for time-consuming amplicon normalization, as shown by the uniform coverage distribution within patients (**Figure 2**). In addition, the use of

Table 2 Overview of mutations and unclassified variants identified in the mutation-negative patients

Patient	Gene	cDNA	Protein	Consolidation			Error correction			Missense predictions			
				Freq	Cov	Score	Freq	Cov	Score	Grantham	PolyPhen	SIFT	Other
7	GUCY2D	c.2302C>T	p.Arg768Trp	58	72	30	48	3,190	29	101	Probably damaging	Deleterious	rs61750168 ⁶⁵
		c.2182G>A	p.Asp728Asn	52	33	27	49	1,235	26	23	Probably damaging	Deleterious	
Lane 1	AIPL1	c.140C>G	p.Thr47Arg	50	22	29	43	2,014	26	71	Probably damaging	Tolerated	UV
		c.937G>T	p.Alala313Ser	50	34	27	45	746	28	99	Benign	Tolerated	rs115681466
1	GUCY2D	c.2577G>T	p.=	53	89	28	54	551	27	/	/	/	rs112372281, Splicing defect?
		c.176T>G	p.Leu59Arg	100	29	28	100	3,237	27	102	Probably damaging	Deleterious	
13	GUCY2D	c.2577G>T	p.=	48	27	20	46	767	20	/	/	/	rs112372281, Splicing defect?
		c.885del	p.Ser296ProfsX10	100	28	23	93	850	22	/	/	/	Frameshift
21	MERTK	c.2237A>G	p.Lys746Arg	52	27	28	51	4,032	28	26	Probably damaging	Deleterious	
		c.1893C>G	p.Ile631Met	55	22	24	49	1,156	24	10	Probably damaging	Deleterious	
14	IQCB1	c.1441G>A	p.Glu481Lys	52	25	28	49	3,239	28	56	Probably damaging	Deleterious	
		c.3773C>T	p.Thr1258Ile	54	41	28	54	5,037	28	89	Possibly damaging	Tolerated	UV
20	CEP290	c.461_463del	p.Glu154del	36	66	20	31	1,908	24	/	/	/	UV

Variants shown in bold type are novel. The Grantham score refers to the physicochemical distance between the two amino acids. The Grantham distance as well as PolyPhen and SIFT predictions were calculated using the Alamut mutation interpretation software v1.54 (Interactive Bioinformatics) (see references therein).

/, not applicable; cDNA, complementary DNA; Cov, coverage; Freq, frequency; SIFT, sorting intolerant from tolerant; p.=, silent variant, no protein change; UV, unclassified variant.

a single qPCR protocol enabled high-throughput amplification in a small volume, thereby overcoming the need for complex PCR multiplexing (Figure 1).³¹ Moreover, qPCR enrichment provides an additional internal quality control. The majority of the 18 failed amplicons (5%) were characterized by secondary structures and a high GC content (Supplementary Table S4 online). The additional optimization required will, however, be more straightforward and flexible in comparison with extra design rounds necessary in case of hybridization-based capturing.²⁰ In general, the amplification of GC-rich regions could be improved by the addition of qPCR enhancers.³⁵ In addition, base-composition bias during the Illumina library preparation has recently been addressed.³⁶ As for now, these 18 amplicons are included in the Sanger confirmation step, which is currently indispensable for the confirmation of variants identified by MPS.

Other differences lie in the sequencing technology. In contrast to resequencing chips,^{30,31} MPS is able to detect small deletions, as illustrated here (Tables 1 and 2 and Supplementary Table S5). The applied workflow of reproducible ligation and shearing enabled sequencing of amplicons with a variable length of up to 650 bp by 1×100 or 2×45 cycles of the Illumina Genome Analyzer IIx. In addition to its large capacity, this platform was chosen because of a lower number of false-positives in homopolymeric regions, characteristic for 454 sequencing.^{22,32} The use of indexing allowed pooling of 12 patients in one Genome Analyzer IIx lane, reducing false-negatives common to untagged pooling strategies.³⁴

Finally, one of the major advantages of our workflow is its flexibility, which is crucial for genetically heterogeneous disorders for which a number of disease genes remain to be identified (i.e., 30% in LCA). This feature is often lacking in both resequencing arrays and hybridization-based capturing. Our workflow enables not only easy addition of new genes but also expansion of regions of interest from coding regions to, for instance, all exons, as applied here. Sequencing of untranslated regions on a large scale has not yet been performed for retinal disease genes and might provide novel insights into the regulation of retinal gene expression.

Overall, the lack of time-consuming optimization of amplification assays, as well as amplicon normalization, the decreasing cost of qPCR with an increasing number of samples to be analyzed, and the use of sample pooling on a short-read sequencer with high capacity, make this approach very cost-effective and thus highly suitable to a clinical context. Of note, the cost and turn-around time of MPS of all LCA genes are lower and shorter, respectively, in comparison with our current routine workflow consisting of LCA chip analysis followed by tedious gene-by-gene Sanger sequencing of a limited gene set.¹⁶

The workflow was subjected to a thorough validation of 107 previously identified variants using NextGENe. This software provides a consolidation method of condensation that merges and elongates reads containing the same anchor sequence. Using this tool, we observed a more evenly distributed coverage (Figure 2) and a lower number of sequencing errors. However,

three heterozygous polymorphisms (2.8%), each present in the original reads, were eliminated (Supplementary Figure S4 online). Importantly, the variants were present following mapping using error-correction condensation, a method that corrects low-frequency errors but does not merge reads. The differences between the two condensation methods stress the importance of dedicated evaluation of data analysis tools for MPS projects. In addition, one should be careful with the exclusion of reported variations from the mutation report, as this could lead to exclusion or incorrect nomenclature of mutations located at the same nucleotide position.

This study included 17 patients for whom the genetic defect was not yet identified following Sanger sequencing of six genes and/or LCA chip analysis. Using the current workflow, the causal genetic defect was found in 3 of 17 patients, in addition to the involvement of two *AIPL1* unclassified variants in a fourth patient (Table 2). Taking into account the contribution of all 16 genes to LCA (~70%)¹ and the detection rate of the LCA chip in the Belgian population (~41%),¹⁶ mutations were expected in 29% or 5/17 individuals. The lower detection rate is presumably due to the additional prescreening of six genes by Sanger sequencing performed in 5 of 17 patients.

Patient 7 is heterozygous for two known *GUCY2D* mutations (p.Arg768Trp and p.Asp728Asn). Of these, p.Arg768Trp was a missed call on the LCA chip, for which ~3.5% of interrogations fail.¹³ Moreover, MPS revealed a second inconsistency of the LCA chip, namely a homozygous instead of heterozygous call for c.907-16_907-14del. Patients 13 and 17 are homozygous for the novel mutations p.Leu59Arg (*RDH12*) and p.Ser296ProfsX10 (*AIPL1*), respectively. Interestingly, patient 13 is also heterozygous for an unclassified variant with potential splicing effect (Table 2). The occurrence of variants in multiple LCA genes has previously been reported.^{12,37} So far, it is unclear whether they influence disease penetrance or represent modifying alleles. Undoubtedly, application of MPS-based strategies on a large scale will reveal more insights into the understanding of intra- and interfamilial variability, individual phenotypes, and a patient's visual prognosis.

In line with this, the MPS panel identified a single heterozygous mutation or unclassified variant in six additional patients (Table 2). First, these might represent modifier alleles influencing mutated alleles of yet to be identified LCA genes. Second, the mutation on the second allele might have been missed, as our current approach is not able to detect deep-intronic mutations, regulatory variants (other than in the 5' and 3' untranslated regions), or large copy number variations. A recently reported 9-kb deletion in *MERTK*, in which we found two novel heterozygous mutations (Table 2), serves as an example of this.³⁸ Ideally, the described MPS approach should be complemented with accurate copy number variation detection, using, for instance, multiplex ligation-dependent probe amplification,³⁹ high-resolution array comparative genomic hybridization,⁴⁰ or qPCR.⁴¹ Of note, qPCR for copy number variation detection should meet specific assay requirements that are not included in the current workflow.⁴¹

The MPS panel presented here is the most comprehensive molecular test currently available for LCA. Identification of a molecular diagnosis is highly important for several reasons. First, it unequivocally confirms the clinical diagnosis, which sometimes offers not only a visual but also a systemic prognosis. Indeed, our panel includes genes associated with LCA as part of a syndrome (*CEP290*, *IQCB1*) in addition to the nonsyndromic LCA genes. The identification of mutations in such genes contributes importantly to general clinical management. Second, a molecular diagnosis opens options for family planning such as prenatal diagnosis or preimplantation genetic diagnosis and provides the basis for recurrence risk assessment. Last but not least, the knowledge of the molecular defect is a prerequisite for gene-specific therapy, as recently established for LCA.^{3–8} This study appointed mutations in *AIPL1*, *GUCY2D*, and *RDH12* as the molecular cause of LCA in three patients. For both *AIPL1* and *GUCY2D*, proof-of-concept studies in animal models have shown beneficial effects of subretinal delivery of adeno-associated vectors containing the wild-type coding sequence.^{42–44} The young age of these patients (10, 2, and 5 years) and the knowledge of their molecular defect identified them as potentially eligible for future gene therapy trials.

In conclusion, we present the first comprehensive molecular test for LCA, based on a novel workflow that allows accurate, fast, flexible, and straightforward enrichment of exons with a variable size range, followed by MPS on a short-read sequencer. The thorough validation and low costs characteristic of our approach argue for its implementation in a clinical context. The strategy used here could easily be applied to other genetically heterogeneous disorders such as other retinal dystrophies, deafness, ataxia, and arrhythmia.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

We thank Sarah De Keulenaer, Jean-Pierre Renard, and Hendrik Van de Voorde for their technical assistance. This study was supported by Research Foundation–Flanders (FWO) grant 31524611 (E.D.B.) and the Fund for Research in Ophthalmology (F.C.). E.D.B. and B.P.L. are senior clinical investigators of the FWO. F.C. and F.P. are postdoctoral fellows of the FWO.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- den Hollander AI, Roepman R, Koenekoop RK, Cremers FP. Leber congenital amaurosis: genes, proteins and disease mechanisms. *Prog Retin Eye Res* 2008;27:391–419.
- Leber T. Uber Retinitis Pigmentosa und angeborene Amaurose. *von Graefes Arch Ophthalmol* 1869;15:25.
- Bainbridge JW, Smith AJ, Barker SS, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* 2008;358:2231–2239.
- Hauswirth WW, Aleman TS, Kaushal S, et al. Treatment of leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: short-term results of a phase I trial. *Hum Gene Ther* 2008;19:979–990.
- Maguire AM, Simonelli F, Pierce EA, et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* 2008;358:2240–2248.
- Maguire AM, High KA, Auricchio A, et al. Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* 2009;374:1597–1605.
- Cideciyan AV, Hauswirth WW, Aleman TS, et al. Human RPE65 gene therapy for Leber congenital amaurosis: persistence of early visual improvements and safety at 1 year. *Hum Gene Ther* 2009;20:999–1004.
- Simonelli F, Maguire AM, Testa F, et al. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol Ther* 2010;18:643–650.
- den Hollander AI, Black A, Bennett J, Cremers FP. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J Clin Invest* 2010;120:3042–3053.
- Coppieters F, Lefever S, Leroy BP, De Baere E. CEP290, a gene with many faces: mutation overview and presentation of CEP290base. *Hum Mutat* 2010;31:1097–1108.
- Otto EA, Loeyes B, Khanna H, et al. Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin. *Nat Genet* 2005;37:282–288.
- Zernant J, Klm M, Dharmaraj S, et al. Genotyping microarray (disease chip) for Leber congenital amaurosis: detection of modifier alleles. *Invest Ophthalmol Vis Sci* 2005;46:3052–3059.
- Henderson RH, Waseem N, Searle R, et al. An assessment of the apex microarray technology in genotyping patients with Leber congenital amaurosis and early-onset severe retinal dystrophy. *Invest Ophthalmol Vis Sci* 2007;48:5684–5689.
- Vallespin E, Cantalapiedra D, Riveiro-Alvarez R, et al. Mutation screening of 299 Spanish families with retinal dystrophies by Leber congenital amaurosis genotyping microarray. *Invest Ophthalmol Vis Sci* 2007;48:5653–5661.
- Yzer S, Leroy BP, De Baere E, et al. Microarray-based mutation detection and phenotypic characterization of patients with Leber congenital amaurosis. *Invest Ophthalmol Vis Sci* 2006;47:1167–1176.
- Coppieters F, Casteels I, Meire F, et al. Genetic screening of LCA in Belgium: predominance of CEP290 and identification of potential modifier alleles in AH1 of CEP290-related phenotypes. *Hum Mutat* 2010;31:E1709–E1766.
- Stone EM. Leber congenital amaurosis – a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. *Am J Ophthalmol* 2007;144:791–811.
- Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009;55:641–658.
- Chou LS, Liu CS, Boese B, Zhang X, Mao R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* 2010;56:62–72.
- Hoppman-Chaney N, Peterson LM, Klee EW, Middha S, Courteau LK, Ferber MJ. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. *Clin Chem* 2010;56:1297–1306.
- Raca G, Jackson C, Warman B, Bair T, Schimmenti LA. Next generation sequencing in research and diagnostics of ocular birth defects. *Mol Genet Metab* 2010;100:184–192.
- Dames S, Durtschi J, Geiersbach K, Stephens J, Voelkerding KV. Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes. *J Biomol Tech* 2010;21:73–80.
- Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 2009;10:R32.
- Lind C, Ferriola D, Mackiewicz K, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Hum Immunol* 2010;71:1033–1042.
- Morgan JE, Carr IM, Sheridan E, et al. Genetic diagnosis of familial breast cancer using clonal sequencing. *Hum Mutat* 2010;31:484–491.
- Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* 1991;220:49–65.
- Reese MG, Eckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–323.

28. Lotery AJ, Namperumalsamy P, Jacobson SG, et al. Mutation analysis of 3 genes in patients with Leber congenital amaurosis. *Arch Ophthalmol* 2000;118:538–543.
29. Swain PK, Chen S, Wang QL, et al. Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. *Neuron* 1997;19:1329–1336.
30. Clark GR, Crowe P, Muszynska D, et al. Development of a diagnostic genetic test for simplex and autosomal recessive retinitis pigmentosa. *Ophthalmology* 2010;117:2169–2177.e3.
31. Booij JC, Bakker A, Kulumbetova J, et al. Simultaneous mutation detection in 90 retinal disease genes in multiple patients using a custom-designed 300-kb retinal resequencing chip. *Ophthalmology* 2011;118:160–167.e1.
32. Bowne SJ, Sullivan LS, Koboldt DC, et al. Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. *Invest Ophthalmol Vis Sci* 2011;52:494–503.
33. Simpson DA, Clark GR, Alexander S, Silvestri G, Willoughby CE. Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. *J Med Genet* 2011;48:145–151.
34. Benaglio P, McGee TL, Capelli LP, Harper S, Berson EL, Rivolta C. Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. *Hum Mutat* 2011;32:E2246–E2258.
35. Horáková H, Polakovicová I, Shaik GM, et al. 1,2-propanediol-trehalose mixture as a potent quantitative real-time PCR enhancer. *BMC Biotechnol* 2011;11:41.
36. Aird D, Ross MG, Chen WS, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 2011;12:R18.
37. Wiszniewski W, Lewis RA, Stockton DW, et al. Potential involvement of more than one locus in trait manifestation for individuals with Leber congenital amaurosis. *Hum Genet* 2011;129:319–327.
38. Mackay DS, Henderson RH, Sergouniotis PI, et al. Novel mutations in MERTK associated with childhood onset rod-cone dystrophy. *Mol Vis* 2010;16:369–377.
39. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002;30:e57.
40. Celestino-Soper PB, Shaw CA, Sanders SJ, et al. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum Mol Genet* 2011;20:4360–4370.
41. D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* 2010;50:262–270.
42. Testa F, Surace EM, Rossi S, et al. Evaluation of Italian patients with leber congenital amaurosis due to AIPL1 mutations highlights the potential applicability of gene therapy. *Invest Ophthalmol Vis Sci* 2011;52:5618–5624.
43. Boye SE, Boye SL, Pang J, et al. Functional and behavioral restoration of vision by gene therapy in the guanylate cyclase-1 (GC1) knockout mouse. *PLoS ONE* 2010;5:e11306.
44. Mihelec M, Pearson RA, Robbie SJ, et al. Long-term preservation of cones and improvement in visual function following gene therapy in a mouse model of leber congenital amaurosis caused by guanylate cyclase-1 deficiency. *Hum Gene Ther* 2011;22:1179–1190.
45. Hanein S, Perrault I, Gerber S, et al. Leber congenital amaurosis: comprehensive survey of the genetic heterogeneity, refinement of the clinical definition, and genotype-phenotype correlations as a strategy for molecular diagnosis. *Hum Mutat* 2004;23:306–317.