A Survey of DNA Variation of *C2ORF71* in Probands with Progressive Autosomal Recessive Retinal Degeneration and Controls

Panagiotis I. Sergouniotis,[1,2] *Zheng Li,*[*,1,3] *Donna S. Mackay,*[1] *Genevieve A. Wright,*[2] *Arundhati Dev Borman,*[1,2] *Sophie R. Devery,*[2] *Anthony T. Moore,*[1,2] *and Andrew R. Webster*[*,1,2]

Purpose. Mutations of *C2ORF71* have recently been reported to be associated with autosomal recessive (AR) retinitis pigmentosa (RP) in humans and with visual defects in zebrafish. *C2ORF71* is located on 2p23.2 and encodes a 1288-amino-acid protein of unknown function, predominately expressed in the photoreceptors. The study was conducted to determine the prevalence of mutations in *C2ORF71* in a cohort of probands with AR retinal degeneration and to detect coding sequence variation in controls.

Methods. A combination of high-resolution DNA melting (HRM) analysis and automated DNA sequencing was used to screen for *C2ORF71* in 286 affected unrelated individuals. Among them, 95 subjects had Leber congenital amaurosis, and 191 had AR RP. In a similar fashion, 151 European and 40 South Asian control DNAs were screened.

Results. Overall, 40 DNA sequence variants were detected, with 17 novel polymorphisms found in the control subjects (8 missense, 7 synonymous, and 2 other). Importantly, 11 novel sequence variants (6 missense and 5 synonymous) in 20 alleles were detected in the cohort of patients but not in the controls. Only one proband was a compound heterozygote but segregation analysis revealed her unaffected father to be homozygous for one of the putative mutations.

Conclusions. *C2ORF71* is a highly polymorphic gene (average heterozygosity of coding region in controls: 2.118×10^{-3}) with many rare variants that confound mutation detection. Further analysis will determine the spectrum of retinal disease caused by mutations in *C2ORF71* and distinguish true pathogenic alleles from the high background of polymorphism elucidating the role of this rare cause of RP in the visual process. (*Invest Ophthalmol Vis Sci.* 2011; 52:1880–1886) DOI:10.1167/iovs.10-6043

From the [1]University College London (UCL), Institute of Ophthalmology, London, United Kingdom; [2]Moorfields Eye Hospital, London, United Kingdom; and the [3]Department of Ophthalmology, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, Peoples Republic of China.

Supported by the British Retinitis Pigmentosa Society, Fight for Sight, Alexander S. Onassis Public Benefit Foundation, Moorfields Eye Hospital Special Trustees, National Institute for Health Research UK (Moorfields Eye Hospital and Institute of Ophthalmology), and the Foundation Fighting Blindness (United States).

Submitted for publication June 11, 2010; revised August 2, 2010; accepted August 12, 2010.

Disclosure: **P.I. Sergouniotis**, None; **Z. Li**, None; **D.S. Mackay**, None; **G.A. Wright**, None; **A. Dev Borman**, None; **S.R. Devery**, None; **A.T. Moore**, None; **A.R. Webster**, None

*Each of the following is a corresponding author: Andrew R. Webster, UCL, Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK; andrew.webster@ucl.ac.uk.

Zheng Li, Department of Ophthalmology, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, P. R. China; zheng.li@ucl.ac.uk.

Retinitis pigmentosa (RP; MIM 268000) denotes a group of genetically determined retinal dystrophies exhibiting genetic and clinical heterogeneity and resulting in photoreceptor cell death. Inheritance may be autosomal dominant (AD), autosomal recessive (AR), X-linked (XL), or digenic. The most prevalent and genetically heterogeneous type is AR, with more than 25 genes implicated so far (RetNet [Retinal Information Network]).

In 2009 Baye et al.,[1] followed by Nishimura et al.[2] and Collin et al.,[3] used homozygosity mapping to identify missense and truncating mutations in a novel gene called *C2ORF71* in five families of various ethnic origins (including Arab Israeli, Spanish, and Dutch) affected with AR retinal dystrophy. Most affected individuals had adult-onset retinal disease,[2,3] but in one family, there was considerable intrafamilial variability, with both childhood- and adult-onset disease.[2]

Human *C2ORF71* is a two-exon gene spanning a 12.5-kb region on 2p23.2. It is conserved in vertebrates, but has no evident paralogues. The gene encodes a 1288-amino-acid photoreceptor-specific protein. Subcellular localization is hypothesized to be within the outer segment and/or the connecting cilium of the photoreceptor cells.[2] The function of *C2ORF71* remains unknown. It does not contain any known functional domains, but does include motifs that suggest posttranslational lipid modification.[2] From the expression patterns in the developing mouse eye, *C2ORF71* is presumed to play a role in retinal development.[3]

High-resolution melting (HRM) is a rapid and cost-effective method of targeted genotyping and mutation screening.[4] It is based on the analysis of the melting profile of amplified DNA and uses fluorescent dye to monitor the transition from unmelted to melted polynucleotide. The approach was first introduced in 1997,[5,6] and recent advances in instrumentation and DNA-binding dyes, have increased its sensitivity, accuracy, and clinical utility.

In this study, we combined HRM with PCR and Sanger sequencing to perform genetic variation analysis of *C2ORF71* in cohorts of probands with various forms of AR retinal degeneration and controls. We sought to determine the prevalence of mutations and polymorphisms in *C2ORF71* and to interpret coding sequence variation. To quantify the degree of polymorphism of the coding sequence in our control population, we calculated the mean heterozygosity and compared it with that of three other retinal disease genes examined in a similar fashion.[7]

Methods

Study Subjects

Two hundred eighty-six affected unrelated individuals ascertained from the clinics of Moorfields Eye Hospital were evaluated for DNA

Investigative Ophthalmology & Visual Science, March 2011, Vol. 52, No. 3

variants. Ninety-five of these were affected with recessive Leber congenital amaurosis (LCA) or childhood-onset retinal dystrophy. The panel was enriched for mutations in novel genes, as the patients had been screened and excluded for known LCA variants through an APEX (arrayed primer extension) genotyping microarray (LCA Chip; Asper Biotech, Tartu, Estonia). The remaining 191 individuals had a diagnosis of AR adult-onset rod–cone or cone–rod dystrophy with unknown molecular diagnosis. The DNA of 151 European (Human Random Control DNA Panels, Salisbury, UK) and 40 South Asian control individuals with normal vision were also screened. Parental DNA was tested to determine the phase of interesting variants on one occasion. Informed consent was obtained from all participants. The study was approved by the local ethics committee and adhered to the tenets of the Declaration of Helsinki.

Genetic Analysis

Genomic DNA was extracted from peripheral blood lymphocytes of the donated blood samples. Amplimers were designed to cover the coding region and intron–exon boundaries of the two published exons (NM_001029883.1). Primers were designed to avoid the known polymorphisms, and their properties were evaluated with OligoCalc[8] (primer sequences and PCR conditions are available on request). The resulting amplimers ranged from 322 to 598 base pairs. Two methods of variation screening were used. On the basis of the reported polymorphisms in *C2ORF71* at the time of the experiment design (18 polymorphic sites, NCBI dbSNP database, November 2009), the amplimers were divided in two groups:

Four of 10 (group 1), covering two or more polymorphic sites each, were amplified by PCR, and mutation analysis was performed by direct sequencing of all PCR products.

Six of 10 amplimers (group 2), covering one or no polymorphic sites each, were analyzed by PCR (with highly saturating fluorescent dye), HRM analysis, and Sanger sequencing.

An in-scale schematic of the *C2ORF71* coding region, showing the coverage by either HRM or direct sequencing and the polymorphism distribution before and after the study, is presented in Table 1.

The decision to perform direct sequencing on group 1 and HRM on group 2 was based on the assumption that three or more common polymorphic sites in an amplimer would impede the analysis of HRM curves (for example, small differences in fluorescence would not be highlighted).[9]

Group 1: PCR and Sanger Sequencing. PCR was performed in a 30 μL volume containing 1× reaction buffer (with 1.5 mM Mg^{2+}) and 0.1× enhancer (Molzym, Bremen, Germany), 0.2 mM of each dNTP, 0.5 μM of each forward and reverse primer, 10 to 50 ng gDNA, and 1 U *Taq* polymerase (MolTaq; Molzym). The enzyme error frequencies for misincorporation and frameshift mutations are 10^{-4} and 2×10^{-5}, respectively (according to the manufacturer's instructions). The thermal cycling profile was initial denaturation (94°C for 3 minutes), amplification (34 cycles of: 94°C denaturation for 30 seconds, 60°C annealing for 30 seconds, and 72°C extension for 30 seconds), and final extension (72°C for 5 minutes).

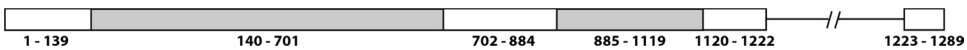
Amplicons were run on a 2% (weight/volume) low-melting-temperature agarose gel in 1× Tris-acetate-EDTA (TAE) buffer to check for the quality and specificity of the PCR reaction. The products were purified with multiscreen PCR filter plates (Millipore, Billerica, MA) before sequencing. Sanger sequencing of purified PCR products was performed with fluorescence-labeled dideoxynucleotides (BigDye Terminator ver. 3.1 Cycle Sequencing Kit; Applied Biosystems Inc. [ABI], Foster City, CA), and sequenced fragments were analyzed on a genetic analyzer (model 3730; ABI). Sequencing-analysis software (Lasergene, ver. 8.02; DNASTAR Inc., Madison, WI) was used to compare the generated electropherograms with those from control samples.

Group 2: PCR, HRM, and Sanger Sequencing. The PCR reactions were performed in a final reaction volume of 10 μL, using 1× reaction buffer, 0.2× enhancer (Molzym), 0.125 mM of each dNTP, 1 μM of forward and reverse primer, 10 to 50 ng of genomic DNA, 10 U of *Taq* polymerase (MolTaq; Molzym), and 1× saturating DNA dye (LCGreen Plus; Idaho Technology, Inc., Salt Lake City, UT). Each reagent concentration was titrated and optimized by comparing the 2% (weight/volume) agarose gel band appearance of control DNA amplified with our method to the one amplified with 2.5× mastermix (LightScanner High Sensitivity Master Mix; Idaho Technology) according to the manufacturer's instructions. Fifteen microliters of mineral oil (Sigma-Aldrich, St. Louis, MO) were added to each well before the reaction commenced, to avoid evaporation. The PCR was initiated with a 5-minute denaturation at 94°C, followed by amplification (40 cycles of: 94°C denaturation for 30 seconds, 67.5–70°C annealing for 30 seconds, and 72°C extension for 1 minute) and final extension (72°C for 7 minutes).

After PCR, HRM was performed in 96-well plates. An HRM genotyping system (LightScanner; Idaho Technology) was used to collect data from 77°C to 95°C at a ramp rate of 0.10°C per second. Images of DNA melting were captured by CCD (charge-coupled device) camera

TABLE 1. (**A**) Coverage of *C20RF71* Coding Sequence by HRM (gray) or Direct Sequencing (white); (**B**) Amplimer Distribution, Number of Variants in dbSNP at the Time of the Experiment Design

A. *C20RF71* Coding Sequence



B. Amplimer Distribution

Amplimer (Amino Acid Positions of *C20RF71* Covered)	Number of Variants: In dbSNP at Time of Experiment Design*	Number of Variants: Identified in 191 Control DNAs
1.1 (1–139)	5	3
1.2, 1.3, 1.4, 1.5 (140–701)	2	6
1.6 (702–884)	5	6
1.7, 1.8 (885–1119)	1	7
1.9 (1120–1222)	2	3
2 (1223–1289)	3	3

* November 2009.

and magnified. Sample-to-sample comparisons of these images were used to interrogate the sequences of the amplified DNA.

Interpretations of data were performed with software allied to the HRM system (LightScanner Instrument and Analysis software with Call-IT function, version 2.0; Idaho Technology). After the negative control was defined, the raw melting data of different samples were normalized; lower temperature ranged between 83°C and 93°C, and upper temperature ranged between 90°C and 95°C. The default value of 5% was chosen for the melting-temperature curve shift function. Samples were clustered into groups by using various curve shape–matching algorithms. Different sensitivity levels were tried (all >0.25), and both normalized melting curve and difference plots were inspected. Samples with significant differences in fluorescence were selected, purified, sequenced bidirectionally, and analyzed as described above. The analysis was possible because the post-PCR product remains intact after processing with the LightScanner system, enabling downstream analysis. To avoid false-positive results, we confirmed the DNA variants identified in fewer than three control or patient DNA samples by using independent regular PCR followed by direct sequencing.

Genetic Variation Quantification

As a large amount of DNA sequence alterations were being identified, there was an interest in quantifying the coding sequence variation and allelic complexity of *C2ORF71* and comparing it with other genes. We used the total expected amount of heterozygosity and its number density per nucleotide as measures of genetic variation. Expected heterozygosity (b_E) is defined as the probability that an individual is heterozygous at a site. The probability of heterozygosity at a site equals 1 minus the probability of homozygosity for each allele. The individual probabilities from m randomly associated sites can be algebraically summed to give a total value of heterozygosity in a sample. Its value assuming Hardy-Weinberg equilibrium is:

$$\sum_{j=1}^{m}(1 - \sum_{i=1}^{n} p_{ij}^{2})$$

where p_{ij} denotes the prevalence of the ith of n alleles (maximum of two alleles per site in our case) at the jth of m sites. The total heterozygosity density per nucleotide is equal to the average heterozygosity, $\hat{H}$ as defined by Nei and Roychoudhury[10]:

$$\hat{H} = \sum_{j=1}^{m} b_j/m = \sum_{j=1}^{m}(1 - \sum_{i=1}^{n} p_{ij}^{2})/m$$

where b_j denotes the expected heterozygosity at the jth of m sites.

On the assumption of the standard neutral model and taking into account the many sites tested, the low average minor allele frequency of the detected polymorphisms, and the large number of control subjects, $\hat{H}$ of a sample of controls estimates the expected average heterozygosity in a population and is a good measure of genetic variability—comparable to nucleotide diversity (π)[11,12] and Watterson's estimate (θ_w).[13]

Statistical Analysis

An individual χ^2 test was performed for each variant with high heterozygote carrier frequency to check the goodness of fit with the Hardy-Weinberg equilibrium (1 *df*).

Web Resources

RetNet, http://www.sph.uth.tmc.edu/Retnet/ University of Texas Houston Health Science Center, Houston, TX

Ensembl, http://www.ensembl.org/[14]

Uniprot, http://www.uniprot.org/[15]

NCBI dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/ National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD

Eukaryotic Linear Motif (ELM), http://elm.eu.org/ The ELM Consortium

SMART, http://smart.embl-heidelberg.de/ European Molecular Biology Laboratory, Heidelberg, Germany

InterPro, http://www.ebi.ac.uk/Tools/InterProScan/ European Bioinformatics Institute, European Molecular Biology Laboratory, Heidelberg, Germany

CDD, http://www.ncbi.nlm.nih.gov/structure/cdd/cdd.shtml National Center for Biotechnology Information

SIFT, http://sift.jcvi.org/[16]

Polyphen, http://genetics.bwh.harvard.edu/pph/[17]

Polyphen 2, http://genetics.bwh.harvard.edu/pph2/[18]

Clustlw, http://www.ebi.ac.uk/Tools/clustalw2/ European Bioinformatics Institute, European Molecular Biology Laboratory, Heidelberg, Germany

GlobProt2, http://globplot.embl.de/[19]

Results

In this study, we identified a total of 40 DNA sequence variants, 28 of which were novel. Among the novel changes, 17 (8 missense, 7 synonymous, one 3-bp deletion, and one 2-bp insertion–deletion) were detected in the controls. The remaining 11 (6 missense and 5 synonymous) were found in 20 alleles of 19 affected subjects (3.5% of all patient alleles) and were not detected in control DNAs. No truncating variants were identified. Two missense changes, c.530C>G, p.Pro177Leu and c.679G>A, p.Glu227Lys, were detected in one AR RP patient in trans. Segregation analysis revealed her unaffected father to be homozygous for c.679G>A, p.Glu227Lys (Fig. 1B). Multiple primer pairs were used to exclude allele-specific amplification due to a variant underlying the primer. This patient was the only one in whom we detected two variants that were not identified in control DNAs. The minor allele frequency in many of the variants found in control populations precluded them from being causative of retinal disease. A c.3248C>T, p.Pro1083Leu variant was detected by HRM in two LCA patients and not in the controls, but the change could not be confirmed with regular PCR and sequencing in either (Fig. 2). All DNA variation in *C2ORF71* identified in our cohort of 286 patients and 191 controls is summarized in Table 2.

The proportion of polymorphic sites in *C2ORF71* was 5% (64/1288). The mean expected heterozygosity of the coding region in unaffected subjects was 2.728 (value for affected samples: 2.48). This means that, on average, 2.728 nucleotides differ in the *C2ORF71* sequence of two randomly chosen chromosomes of the population. The probability that each nucleotide would be nonidentical between two randomly chosen control subject sequences is approximately the average heterozygosity $\hat{H}$ and in *C2ORF71* equals 21.18×10^{-4} (19.26×10^{-4} in affected samples). This means that if two chromosomes have been chosen at random, a sequence variation would have been identified every 472 bp. These values are higher than those of the *BEST1* (0.14 sites, 0.8×10^{-4} sites/nucleotide), *EFEMP1* (0.003 sites, 0.03×10^{-4} sites/nucleotide), and *ABCA4* (1.28 sites, 1.8×10^{-4} sites/nucleotide) genes in unaffected subjects,[7] showing that the *C2ORF71* coding sequence is more variable in the control population. For human genomic DNA, average heterozygosity is on the order of 8×10^{-4}, approximately 1 of 1250 bp,[20,21] with

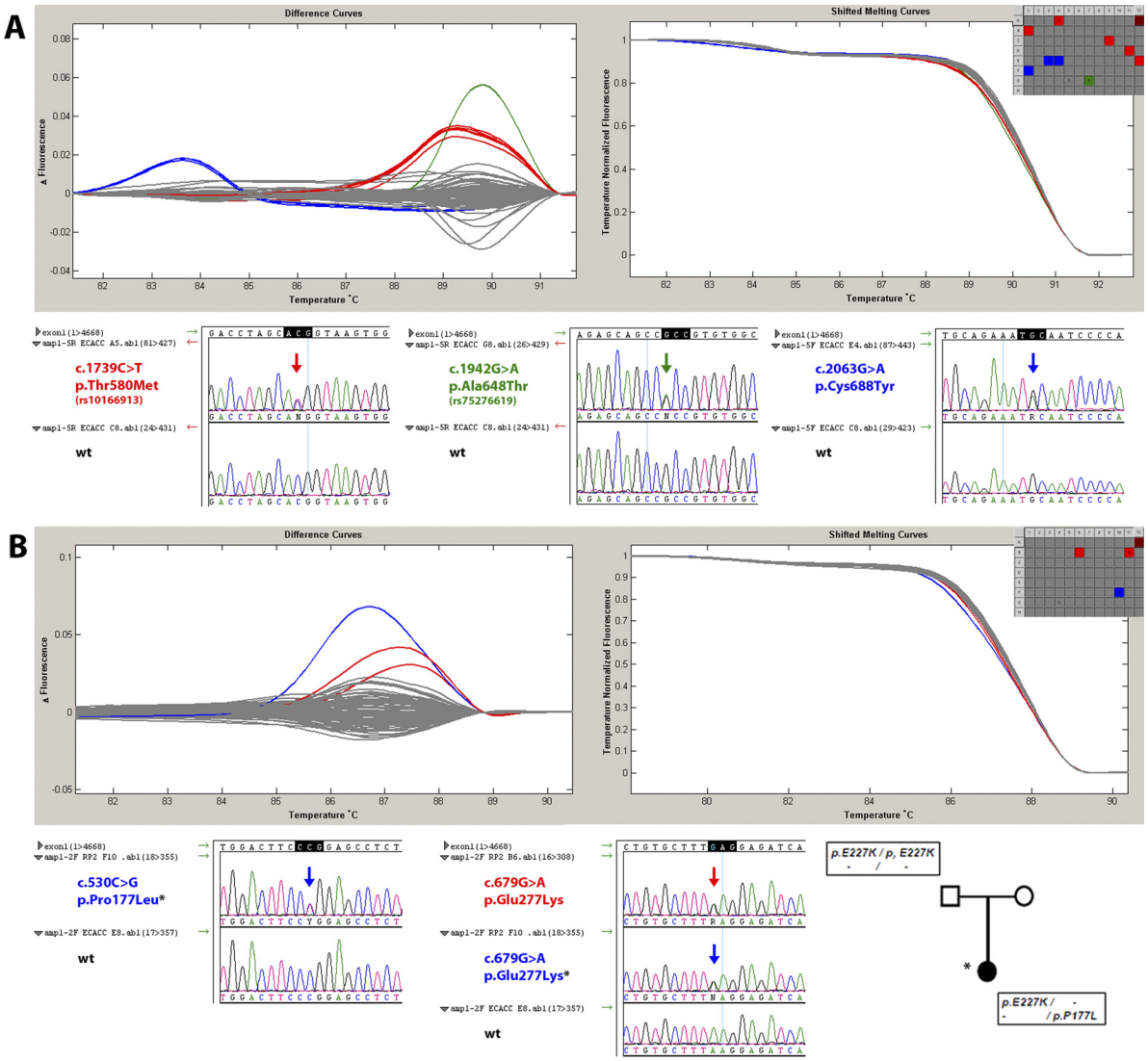


FIGURE 1.

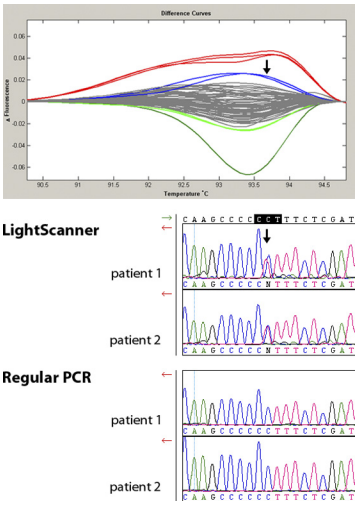


FIGURE 2.

FIGURE 1. HRM normalized curve and difference plots and electropherograms generated by downstream sequencing of the amplified product. (**A**) Amplimer 1.5 for European control plate 1: Two clusters of samples (*blue* and *red* groups) and one isolated sample (*green* group) demonstrate melting difference plots that significantly deviated from most of the control DNAs screened in the plate. When sequenced, samples from the same group contained the same missense change in the heterozygous state (c.1739C>T, p.Thr580Met for the *red* group; c.2063G>A, p.Cys688Tyr for the *blue* group; and c.1942G>A, p.Alal648Thr. for the *green* group). (**B**) Amplimer 1.2 for RP plate 2: Three samples demonstrated differences in fluorescence compared with most of the examined samples. Sequencing revealed that two samples were heterozygous for c.679G>A, p.Glu277Lys (*red* group), and one was compound heterozygous for c.530C>G, p.Pro177Leu and c.679G>A, p.Glu277Lys (*blue* group). Neither of those two changes was identified in our control DNA panels. Family members of the patient in the *blue* group were screened, with the unaffected father being homozygous for c.679G>A, p.Glu277Lys.

FIGURE 2. HRM curve difference plots for amplimer 1.8 in an LCA panel are presented. Among the different groups, two samples with similar melting behavior can be observed (*blue* group). Electropherograms of amplimer 1.8 in those two patient samples were generated. *Rows 1* and *2* are the result of sequencing the product of amplification using LightScanner protocol, while *rows 3* and *4* show the same samples amplified through normal PCR. A c.3248C>T, p.Pro1083Leu mutation was suspected initially, but could not be confirmed in either patient.

TABLE 2. Summary of *C20RF71* Sequence Variants; Predicted Impact on C2ORF71 and Frequency in Patients and Controls

Nucleotide	DNA Variants Protein	dbSNP	SIFT* Prediction	Tolerance Index (0–1)	Polyphen†-Polyphen 2‡ Prediction§ Ver1†	Ver2‡	PSIC Score Difference (0–4)†	Hum Div Score (0–1)‡	Frequency Controls	Patients
c.37A>T	p.Ser13Cys	rs10084168	Intolerant	0.01	POS		1.548	0.774	0/364	1/572
c.60G>A	p.Gln20Gln	rs35929540			N/A				82/364	101/572
c.99G>A	p.Gln33Gln				N/A				1/364	2/572
c.158 C>T	**p.Ala53Val**		Tolerant	0.26	Benign		1.129	0.022	0/364	1/572
c.258G>A	p.Arg86Arg	rs62132765			N/A				77/364	111/572
c.530C>T	**p.Pro177Leu**		Intolerant	0.00	PRB		2.550	0.999	0/382	1/572
c.531G>A	**p.Pro177Pro**				N/A				0/382	1/572
c.679G>A	**p.Glu227Lys**		Intolerant	0.01	Benign	PRB	1.409	0.949	0/382	3/572
c.1262A>G	p.Lys421Arg	rs17007544	Tolerant	0.47	Benign	POS	1.047	0.665	3/190	0/572
c.1387G>T	p.Val463Phe		Tolerant	0.05	Benign	POS	1.284	0.375	1/190	1/572
c.1452C>T	p.Ser484Ser	rs13385188			N/A				81/190	153/572
c.1739C>T	p.Thr580Met	rs10166913	Intolerant	0.01	POS		1.665	0.798	5/190	0/572
c.1844T>A	**p.Val615Asp**		Tolerant	0.40	POS	PRB	1.970	0.925	0/190	1/572
c.1942G>A	p.Ala648Thr	rs75276619	Tolerant	0.59	Benign		0.211	0.022	1/190	11/572
c.2063G>A	p.Cys688Tyr		Tolerant	0.52	PRB		2.938	0.996	3/190	0/572
c.2112T>C	p.Asn704Asn	rs10200693			N/A				123/380	168/572
c.2234G>A	p.Arg745Lys		Tolerant	0.92	Benign	POS	1.132	0.788	1/380	0/572
c.2374C>G	p.Leu792Val	rs17744093	Tolerant	0.30	Benign	POS	1.017	0.426	66/380	103/572
c.2406A>G	p.Ala802Ala				N/A				1/380	1/572
c.2498C>T	**p.Pro833Leu**		Intolerant	0.01	PRB		2.550	0.994	0/380	1/572
c.2499G>A	p.Pro833Pro	rs34253433			N/A				70/380	106/572
c.2502T>C	**p.Pro834Pro**				N/A				0/382	1/572
c.2600C>T	p.Pro867Leu		Tolerant	0.72	Benign		0.213	0.20	1/380	2/572
c.2864G>A	p.Arg955Gln		Tolerant	0.81	Benign		0.192	0.192	1/190	0/572
c.2875G>A	p.Ala959Thr		Tolerant	0.23	Benign		0.995	0.109	2/190	1/572
c.2889C>T	p.Ser963Ser				N/A				2/190	3/572
c.3037C>G	p.Pro1013Ala		Tolerant	0.23	PRB		2.100	0.908	1/372	0/572
c.3058_3059del CAinsAG	p.Gln1020Arg		Tolerant	0.52	Benign	N/A	0.900	N/A	2/372	3/572
c.3264_3266del CCC	p.Pro1089del				N/A				1/372	6/572
c.3291G>A	p.Gln1097Gln				N/A				1/372	3/572
c.3395A>C	p.Glu1132Ala	rs78874550	Tolerant	0.32	Benign	N/A	1.350	N/A	1/174	1/572
c.3447G>A	p.Pro1149Pro				N/A				1/174	5/572
c.3522C>T	p.Asp1174Asp				N/A				1/174	2/572
c.3609G>A	**p.Pro1203Pro**				N/A				0/174	1/572
c.3626A>G	**p.Asp1209Gly**		Tolerant	0.43	POS	N/A	1.575	N/A	0/174	1/572
c.3673_3675dup3		rs72183347, rs72122505			N/A				43/300	84/572
c.3739G>A	p.Gly1247Ser		Tolerant	1.00	Benign	N/A	0.225	N/A	3/300	4/572
c.3780C>T	**p. Gly1260Gly**				N/A				0/300	1/572
c.3789G>A	p.Leu1263Leu				N/A				1/300	4/572
c.3840G>A	**p.Ala1280Ala**				N/A				0/300	2/572

* SIFT (Sorting Intolerant from Tolerant)[16]; tolerance index, a lower score indicating intolerance to the substitution.
† Polyphen (Polymorphism Phenotyping)[17]: PSIC (position-specific independent counts). A high difference indicates intolerance.
‡ Polyphen-2 (polymorphism phenotyping v2)[18]: HumDiv Score. A high score indicates intolerance to substitution.
§ PRB, probably damaging; POS, possibly damaging. Novel changes detected only in patients are in bold.

HLA genes being examples of exceptionally polymorphic loci.[22,23]

Discussion

To improve molecular testing of retinal genes in panels of affected individuals and controls we have exploited the advances of HRM technology. We recruited 286 individuals affected with recessive retinal dystrophy and screened them for mutations in *C2ORF71*, a gene recently associated with AR retinal dystrophy. Previously, in our laboratory we used Sanger sequencing as the method of choice for mutation screening. HRM analysis with saturating fluorescent DNA-binding dye is a highly sensitive and specific method for sequence-variation scanning. To date, more than 60 genes have been analyzed by this method, including *ABCA4.*[24,25]

HRM analysis allows DNA sequence variation scanning in two steps: amplification (with a fluorescent dye) and amplicon melting. The melting curve analysis run requires less than 5 minutes and, combined with the minimal additional cost of the dye (LCGreen Plus; Idaho Technology, Inc.), makes LightScanner a fast and cost-efficient method of screening large panels of patients and controls. We chose to use this method for the parts of the *C2ORF71* sequence with one or no reported polymorphisms. We used Sanger sequencing in the remaining amplicons, as the many sequence variants already reported would potentially hinder and complex the analysis and interpretation of the melting curves. Based on our observations, three or more common polymorphic sites in an amplimer contribute negatively to the analysis.

HRM has been markedly efficient in detecting heterozygous variants and is widely used as a method to detect carriers in the *BRCA1* and *BRCA2* genes.[26,27] For heterozygous variants, sensitivity and specificity approach 100% for all categories of substitution, as well as insertions and deletions small enough to be amplified by PCR (reviewed by Taylor[28]). Such variants are detectable at any location in the PCR amplimer, including those within a few base pairs of the primers.[29] Homozygous changes are harder to distinguish from wild-type with 16% of all single-nucleotide polymorphisms presenting minimal or no differences between the melting temperatures of the two homozygous states.[24] Compared with the wild-type, most homozygous sequence changes produce a melting temperature shift and more sharp and symmetric melting transitions,[30,31] whereas heterozygous samples are identified by differences in melting curve shape (distortion) with a more gradual, complex transition.[4,32,33] In this study, only one sequence alteration (c.1452C>T, p.Ser484Ser) was detected in homozygous phase in the *C2ORF71* amplimers analyzed with by HRM. The low carrier frequency of the remaining 16 heterozygous changes could explain the absence of homozygotes (χ^2 goodness of fit test, $P < 0.05$). Despite the fact that the average coefficient of inbreeding in our cohorts is likely to be low, we cannot exclude the possibility that a homozygous change could not be detected with our method.

Common variants can be recognized by characteristic melting patterns[34] although this notion is debatable.[28] We identified common polymorphisms by melting curve identity and rare sequence variants by sequencing. On average, 12% (11 samples) of each 96-well plate was sequenced, although this result varied widely among the amplimers and runs. Downstream sequencing was generated by HRM false-positive results. Importantly, on one occasion, a false-positive result could not be resolved by sequencing directly after the HRM analysis (Fig. 2). Independent regular PCR followed by sequencing revealed that both patients, initially tagged as heterozygotes for a c.3248C>T, p.Pro1083Leu mutation, had a normal sequence. We assume that this substitution may have occurred due to a hot spot for polymerization errors in the sequence of the target DNA and, possibly, to the increased number of cycles our HRM amplification protocol uses compared with normal PCR.[35] This observation made us modify our study design and independently amplify and sequence all variants identified in three or fewer patients or control samples.

During the study, it became evident that the total number of *C2ORF71* sequence variants discovered in controls or reported in dbSNP (48 in total) was disproportionately greater than those in other retinal genes, even without accounting for gene size. However, the proportion of polymorphic nucleotides does not provide an assessment of how variable the coding sequence of the gene is in a population. This fact is more obvious in C2ORF71, as most sequence alterations were found to have small minor allele frequencies. Average heterozygosity $\hat{H}$ was used to summarize variation as a function of both the number of polymorphic sites and their frequencies in the population. $\hat{H}$ calculates the probability of nonidentity of two randomly chosen chromosomal sites.[10] *C2ORF71* demonstrated greater genetic variability compared with *BEST1*, *EFEMP1*, and *ABCA4*.[7]

Comparison of the C2ORF71 peptide sequence with its orthologues revealed minimal conservation (Supplementary Fig. S1, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.10-6043/-/DCSupplemental). Multiple islands of conserved sequence are observed between amino acids 173 and 359. Four of five published mutations (p.Gln186X,[3] p.Ile201Phe,[1,2] p.Asn316MetfsX5[3], and p.Trp253X[2]) fall within this 173-359 interval. No polymorphisms in this region were detected in our control DNA panels with the exception of an unaffected parent who was homozygous for a c.679G>A, p.Glu227Lys change. In addition, two missense variants are reported in dbSNP to be located within this interval (c.740T>C, p.Val247Ala, and c.755C>A, p.Ala252Asp), but these were not detected in our cohort. Other regions, completely conserved between species, included the first three amino acids of the protein and amino acids 830 to 834. The significance of complete conservation of glycine at position 2 and cysteine at position 3 (G2/C3), a motif also found in *RP2* that is subject to lipid modification, has been investigated by Nishimura et al.[2] In the 830-834 amino acid region we identified three variants: a known polymorphism, c.2499G>A, p.Pro833Pro (rs34253433); a missense change, c.2498C>T, p.Pro833Leu, in one LCA patient; and a silent change, c.2502T>C, p.Pro834Pro, in another LCA patient, the latter two being in the heterozygous state. No second change was identified in either of those two affected individuals.

C2ORF71 does not harbor any known functional domains, with results from screening with protein analysis tools being negative. The region between amino acids 176 and 331 is predicted to be a potential globular domain by GlobProt 2,[19] but this may be falsely positive because of the conservation observed in this region. A region of compositional bias is described in UniProt (http://www.uniprot.org/uniprot/A6NGG8) with proline being overrepresented within the subsection between amino acids 1013 and 1095. There is little functional knowledge on C-terminal proline-rich domains (PRDs). However, PRDs are implicated in a number of aberrant protein interactions with certain protein interaction domains preferring ligand sequences that are proline rich.[36] Recently, a PRD in a microtubule-associated protein, mainly existing in neurons, has been shown to mediate interactions with actin.[37] The photoreceptor connecting cilium contains clusters of actin[38] and previous work suggests interaction of C2ORF71 with the connecting cilium.[2] Several genes, such as *TULP1*[39] and *MYO7A*,[40] interacting with actin, are also associated with isolated or syndromic RP.

In this study, we used HRM and Sanger sequencing to detect mutations in *C2ORF71* and encountered the technical challenges of screening a polymorphic gene with numerous rare variants. We investigated how to quantify the high degree of genetic variability and sought to understand what the polymorphic site distribution implies. We could not find convincing evidence that biallelic mutation of *C2ORF71* was responsible for the retinal degeneration in any of our 286 families. Further analysis will determine the function of the encoded protein in the photoreceptor cells.

Acknowledgments

The authors thank the family members who participated in the study for their cooperation and help; our colleagues who referred patients to us at MEH; Beverly Scott and Naushin Waseem for technical assistance; and Alice Davidson for help in revising the manuscript.

References

1. Baye LM, Nishimura DY, Searby CC, et al. *Discovery and Functional Analysis of a New Retinitis Pigmentosa Gene, C2orf71.* Presented at The American Society of Human Genetics 59th Annual Meeting. Program Number 195, Honolulu, Hawai; 2009.
2. Nishimura DY, Baye LM, Perveen R, et al. Discovery and functional analysis of a retinitis pigmentosa gene, C2ORF71. *Am J Hum Genet.* 2010;86:686–695.
3. Collin RW, Safieh C, Littink KW, et al. Mutations in C2ORF71 cause autosomal-recessive retinitis pigmentosa. *Am J Hum Genet.* 2010; 86:783–788.
4. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem.* 2003;49:853–860.
5. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. Continuous fluorescence monitoring of rapid cycle DNA amplification. *BioTechniques.* 1997;22:130–131, 134–138.
6. Ririe KM, Rasmussen RP, Wittwer CT. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem.* 1997;245:154–160.
7. Webster AR, Heon E, Lotery AJ, et al. An analysis of allelic variation in the *ABCA4* gene. *Invest Ophthalmol Vis Sci.* 2001;42:1179–1189.
8. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* 2007;35:W43–46.
9. Olsen RK, Dobrowolski SF, Kjeldsen M, et al. High-resolution melting analysis, a simple and effective method for reliable mutation scanning and frequency studies in the ACADVL gene. *J Inherit Metab Dis.* 2010;33:247–260.
10. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. *Genetics.* 1974;76:379–390.
11. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 1979;76:5269–5273.
12. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–595.
13. Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 1975;7:256–276.
14. Hubbard T, Barker D, Birney E, et al. The Ensembl genome database project. *Nucleic Acids Res.* 2002;30:38–41.
15. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2009;37:D169–D174.
16. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4:1073–1081.
17. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 2002;30:3894–3900.
18. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7: 248–249.
19. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 2003;31:3701–3708.
20. Reich DE, Schaffner SF, Daly MJ, et al. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet.* 2002;32:135–142.
21. Przeworski M, Hudson RR, Di Rienzo A. Adjusting the focus on human variation. *Trends Genet.* 2000;16:296–302.
22. Klein J, Takahata N, Ayala FJ. MHC polymorphism and human origins. *Sci Am.* 1993;269:78–83.
23. Strachan TR, Read AP. *Human Molecular Genetics.* 3rd ed. Oxford, UK: Garland Science; 2004.
24. Erali M, Wittwer CT. High resolution melting analysis for gene scanning. *Methods.* 2010;50:250–261.
25. Aguirre-Lamban J, Riveiro-Alvarez R, Garcia-Hoyos M, et al. Comparison of high-resolution melting analysis with denaturing high-performance liquid chromatography for mutation scanning in the *ABCA4* gene. *Invest Ophthalmol Vis Sci.* 2010;51:2615–2619.
26. van der Stoep N, van Paridon CD, Janssens T, et al. Diagnostic guidelines for high-resolution melting curve (HRM) analysis: an interlaboratory validation of BRCA1 mutation scanning using the 96-well LightScanner. *Hum Mutat.* 2009;30:899–909.
27. de Juan I, Esteban E, Palanca S, Barragan E, Bolufer P. High-resolution melting analysis for rapid screening of BRCA1 and BRCA2 Spanish mutations. *Breast Cancer Res Treat.* 2009;115:405–414.
28. Taylor CF. Mutation scanning using high-resolution melting. *Biochem Soc Trans.* 2009;37:433–437.
29. Dobrowolski SF, McKinney JT, Amat di San Filippo C, Giak Sim K, Wilcken B, Longo N. Validation of dye-binding/high-resolution thermal denaturation for the identification of mutations in the SLC22A5 gene. *Hum Mutat.* 2005;25:306–313.
30. Liew M, Pryor R, Palais R, et al. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin Chem.* 2004;50:1156–1164.
31. Palais RA, Liew MA, Wittwer CT. Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. *Anal Biochem.* 2005;346:167–175.
32. Graham R, Liew M, Meadows C, Lyon E, Wittwer CT. Distinguishing different DNA heterozygotes by high-resolution melting. *Clin Chem.* 2005;51:1295–1298.
33. Laurie AD, Smith MP, George PM. Detection of factor VIII gene mutations by high-resolution melting analysis. *Clin Chem.* 2007; 53:2211–2214.
34. Vandersteen JG, Bayrak-Toydemir P, Palais RA, Wittwer CT. Identifying common genetic variants by high-resolution melting. *Clin Chem.* 2007;53:1191–1198.
35. Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl.* 1991;1:17–24.
36. Kay BK, Williamson MP, Sudol M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 2000;14:231–241.
37. He HJ, Wang XS, Pan R, Wang DL, Liu MN, He RQ. The proline-rich domain of tau plays a role in interactions with actin. *BMC Cell Biol.* 2009;10:81.
38. Chaitin MH, Schneider BG, Hall MO, Papermaster DS. Actin in the photoreceptor connecting cilium: immunocytochemical localization to the site of outer segment disk formation. *J Cell Biol.* 1984;99:239–247.
39. Xi Q, Pauer GJ, Marmorstein AD, Crabb JW, Hagstrom SA. Tubby-like protein 1 (TULP1) interacts with F-actin in photoreceptor cells. *Invest Ophthalmol Vis Sci.* 2005;46:4754–4761.
40. Rayment I, Rypniewski WR, Schmidt-Base K, et al. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science.* 1993;261:50–58.