A Survey of DNA Variation of C2ORF71 in Probandareth With Progressive Autosomal Recessive Retinal Degeneration and Controls

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

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, predominantly 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^-5) 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 Ophtalmol Vis Sci. 2011; 52:1880–1886) DOI:10.1167/iovs.10-40415

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., followed by Nishimura et al. and Collin et al. 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, but in one family, there was considerable intrafamilial variability, with both childhood- and adult-onset disease.

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 paralogs. 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. The function of C2ORF71 remains unknown. It does not contain any known functional domains, but does include motifs that suggest posttranslational lipid modification. From the expression patterns in the developing mouse eye, C2ORF71 is presumed to play a role in retinal development.

High-resolution melting (HRM) is a rapid and cost-effective method of targeted genotyping and mutation screening. It is based on the analysis of the melting profile of amplified DNA and uses fluorescent dye to monitor the transition from unmethylated to methylated polynucleotide. The approach was first introduced in 1997 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.

Methods

Study Subjects

Two hundred eighty-six affected unrelated individuals ascertained from the clinics of Moorfields Eye Hospital were evaluated for DNA variation of C2ORF71.
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* (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:

The decision to perform direct sequencing on group 1 and HRM on group 2 was based on the assumption that three or more common variants in an amplimer would impede the analysis of HRM group 2 was based on the assumption that three or more common variants in an amplimer would impede the analysis of HRM

<table>
<thead>
<tr>
<th>A. C2ORF71 Coding Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–139</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Amplimer Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplimer (Amino Acid Positions of C2ORF71 Covered)</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>1.1 (1–139)</td>
</tr>
<tr>
<td>1.2, 1.3, 1.4, 1.5 (140–701)</td>
</tr>
<tr>
<td>1.6 (702–884)</td>
</tr>
<tr>
<td>1.7, 1.8 (885–1119)</td>
</tr>
<tr>
<td>1.9 (1120–1222)</td>
</tr>
<tr>
<td>2 (1223–1289)</td>
</tr>
</tbody>
</table>

* 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 (\( h_e \)) 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:

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

where \( p_{ij} \) denotes the prevalence of the \( i \)th of \( n \) alleles (maximum of two alleles per site in our case) at the \( j \)th of \( m \) sites. The total heterozygosity density per nucleotide is equal to the average heterozygosity, \( \hat{H} \) as defined by Nei and Roychoudhury10:

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

where \( b_j \) denotes the expected heterozygosity at the \( j \)th 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 (\( \pi \))11,12 and Watterson’s estimate (\( \theta_n \)).13

**Statistical Analysis**

An individual \( \chi^2 \) 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/

Uniprot, http://www.uniprot.org/


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


SIFT, http://sift.jcvi.org/

Polyphen, http://genetics.bwh.harvard.edu/pph2/

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

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

GlobProt2, http://globplot.embl.de/

**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 events, 7 synonymous, 1 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.Pro1771Leu 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 \( \times 10^{-4} \) (19.26 \( \times 10^{-4} \) 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 \( \times 10^{-4} \) sites/nucleotide), EFEMP1 (0.003 sites, 0.03 \( \times 10^{-4} \) sites/nucleotide), and ABCA4 (1.28 sites, 1.8 \( \times 10^{-4} \) sites/nucleotide) genes in unaffected subjects,7 showing that the C2ORF71 coding sequence is more variable in the control population. For human genomc DNA, average heterozygosity is on the order of 8 \( \times 10^{-4} \), approximately 1 of 1250 bp,20,21 with
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/H11022G, p.Thr580Met for the red group; c.2063G>A/H11022G, p.Cys688Tyr for the blue group; and c.1942G>A/H11022G, p.Ala648Thr 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

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>DNA Variants</th>
<th>Protein</th>
<th>dbSNP</th>
<th>SIFT Prediction</th>
<th>Tolerance Index (0-1)</th>
<th>Polyphen Polyphen 2‡</th>
<th>PSIC Score Difference (0-4)†</th>
<th>Hum Div Score (0-1)‡</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.37A&gt;T</td>
<td>p.Ser13Cys</td>
<td>rs10084168</td>
<td>Intolerant 0.01</td>
<td>POS</td>
<td>1.548</td>
<td>0.774</td>
<td>Controls: 0/364 Patients: 1/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.60G&gt;A</td>
<td>p.Glu20Glu</td>
<td>N/A</td>
<td>Tolerant 0.26</td>
<td>Benign</td>
<td>1.129</td>
<td>0.022</td>
<td>Controls: 0/364 Patients: 77/111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.99G&gt;A</td>
<td>p.Glu33Glu</td>
<td>N/A</td>
<td>Tolerant 0.47</td>
<td>Benign</td>
<td>1.047</td>
<td>0.665</td>
<td>Controls: 0/364 Patients: 1/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.158C&gt;T</td>
<td>p.Ala53Val</td>
<td>N/A</td>
<td>Tolerant 0.52</td>
<td>Benign</td>
<td>2.938</td>
<td>0.996</td>
<td>Controls: 0/364 Patients: 5/172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.258G&gt;A</td>
<td>p.Arg86Arg</td>
<td>rs62132765</td>
<td>Intolerant 0.01</td>
<td>Benign</td>
<td>1.409</td>
<td>0.494</td>
<td>Controls: 0/364 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.530C&gt;T</td>
<td>p.Pro177Leu</td>
<td>N/A</td>
<td>Tolerant 0.92</td>
<td>Benign</td>
<td>1.132</td>
<td>0.788</td>
<td>Controls: 0/364 Patients: 1/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.531G&gt;A</td>
<td>p.Pro177Lys</td>
<td>N/A</td>
<td>Tolerant 0.30</td>
<td>Benign</td>
<td>1.017</td>
<td>0.426</td>
<td>Controls: 0/364 Patients: 66/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.679G&gt;A</td>
<td>p.Glu227Lys</td>
<td>N/A</td>
<td>Tolerant 0.52</td>
<td>Benign</td>
<td>2.550</td>
<td>0.994</td>
<td>Controls: 1/3/572 Patients: 1/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1262A&gt;G</td>
<td>p.Lys421Arg</td>
<td>rs17007544</td>
<td>Tolerant 0.47</td>
<td>Benign</td>
<td>1.129</td>
<td>0.022</td>
<td>Controls: 0/364 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1387G&gt;T</td>
<td>p.Val615Asp</td>
<td>N/A</td>
<td>Tolerant 0.52</td>
<td>Benign</td>
<td>2.938</td>
<td>0.996</td>
<td>Controls: 0/364 Patients: 5/172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1452C&gt;T</td>
<td>p.Ser484Ser</td>
<td>rs3385188</td>
<td>Intolerant 0.01</td>
<td>Benign</td>
<td>1.409</td>
<td>0.494</td>
<td>Controls: 0/364 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1739G&gt;T</td>
<td>p.Glu580Glu</td>
<td>N/A</td>
<td>Tolerant 0.92</td>
<td>Benign</td>
<td>1.132</td>
<td>0.788</td>
<td>Controls: 0/364 Patients: 1/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.2234G&gt;A</td>
<td>p.Arg745Lys</td>
<td>rs17744093</td>
<td>Intolerant 0.01</td>
<td>Benign</td>
<td>1.409</td>
<td>0.494</td>
<td>Controls: 0/364 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.2499G&gt;A</td>
<td>p.Pro833Pro</td>
<td>rs34253433</td>
<td>Tolerant 0.52</td>
<td>Benign</td>
<td>2.938</td>
<td>0.996</td>
<td>Controls: 0/364 Patients: 5/172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.3264G&gt;A</td>
<td>p.Pro1089Pro</td>
<td>rs3264_3266del</td>
<td>N/A</td>
<td>0.900</td>
<td>2/372</td>
<td>5/72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.3395C&gt;A</td>
<td>p.Glu1132Ala</td>
<td>rs78874550</td>
<td>Tolerant 0.32</td>
<td>Benign</td>
<td>1.550</td>
<td>0.774</td>
<td>Controls: 1/3/572 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.3609G&gt;A</td>
<td>p.Pro203Pro</td>
<td>N/A</td>
<td>Tolerant 0.43</td>
<td>Benign</td>
<td>1.575</td>
<td>0.774</td>
<td>Controls: 1/3/572 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.3739G&gt;A</td>
<td>p.Leu852Leu</td>
<td>N/A</td>
<td>Tolerant 1.00</td>
<td>Benign</td>
<td>0.225</td>
<td>0.774</td>
<td>Controls: 1/3/572 Patients: 3/572</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* SIFT (Sorting Intolerant from Tolerant): tolerance index, a lower score indicating intolerance to the substitution.
† Polyphen (Polymorphism Phenotyping): § PSIC Score Difference (0-4): A high difference indicates intolerance.
‡ Polyphen-2 (polymorphism phenotyping v2): 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.22,25

**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 amplicon 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 Taylor28). Such variants are detectable at any location in the PCR amplimer, including those within a few base pairs of the primers.24 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,29,31 whereas heterozygous samples are identified by differences in melting curve shape (distortion) with a more gradual, complex transition.32–34 In this study, only one sequence alteration (c.1452C>T, p.Ser484Ser) was detected in homozygous phase with results from screening with protein analysis tools being negative. The region between amino acids 176 and 331 is 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 \( H \) was used to summarize variation as a function of both the number of polymorphic sites and their frequencies in the population. \( H \) calculates the probability of nonidentity of two randomly chosen chromosomal sites.10 C2ORF71 demonstrated greater genetic variability compared with BEST1, EFEMP1, and ABCA4.

Comparison of the C2ORF71 peptide sequence with its orthologues revealed minimal conservation (Supplementary Fig. S1, http://www.iovs.orglookup/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.Gln180X,3 p.Ile201Phe,1,2 p.Asn316MetfsX57, and p.Trp253X) 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 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 (rs4254543); 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 homozygous 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 GlobPlot 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 actin38 and previous work suggests interaction of C2ORF71 with the connecting cilium.3 Several genes, such as TULIP39 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


39. Nei M, Roychoudhury AK. Sampling variances of heterozygosity 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.