Diagnosis of Autosomal Dominant Retinitis Pigmentosa by Linkage-Based Exclusion Screening with Multiple Locus-Specific Microsatellite Markers

Hiroaki Kondo,1 Tomoko Tabira,2 Atsushi Mizota,3 Emiko Adachi-Usami,3 Kenji Oshima,1 and Kenshi Hayashi2

PURPOSE. To describe a hierarchical approach for efficient genetic diagnosis of autosomal dominant retinitis pigmentosa (adRP).

METHODS. Forty di-, tri-, or tetra-nucleotide repeats tightly linked to 10 genes known to be responsible for adRP were identified from the human genome sequence and used as markers in multiplex amplification and genotyping, followed by linkage analysis. Discordance of cosegregation of markers and the disease excluded the majority of the examined genes as candidates, and mutation screening for the remaining genes was performed.

RESULTS. With this strategy, examination of an adRP-affected family indicated that 3 of 10 candidate genes segregated concordantly with the disease. Further searches for mutations revealed a novel insertion and deletion in the last exon of a splicing factor gene, PRPF8.

CONCLUSIONS. This systematic approach facilitates the molecular diagnosis of adRP, which is known to have a highly heterogeneous genetic background. (Invest Ophthalmol Vis Sci. 2003; 44:1275–1281) DOI:10.1167/iovs.02-0895

Retinitis pigmentosa (RP) affects approximately 1.5 million people worldwide. Patients with RP experience night blindness, a gradual loss of peripheral visual field, and eventually loss of central vision, in most cases due to degeneration of the photoreceptor cells of the retina. The defective genes in RP are strikingly heterogeneous. There are autosomal dominant (adRP), autosomal recessive (arRP), X-linked (xlRP), and rare mitochondrial and digenic forms. To date, 13 genes for adRP were cloned shortly after the survey, with the number of genes to be screened increasing sevenfold.11–17 To facilitate the molecular diagnosis for such highly heterogeneous disorders, a systematic approach is necessary.

Genetic linkage analysis has offered a substantial role in identifying novel genes responsible for genetic disease. Even for diseases with known genes, linkage analysis is an essential tool for molecular diagnosis when more than two loci are candidates. However, linkage analysis has an intrinsic weakness, because the possibility of meiotic crossover between gene loci and the available genetic markers remains. Furthermore, analyzing multiple gene loci requires a considerable amount of time and laboratory resources. Nonetheless, we believe that examining the genomic information in the vicinity of the RP-causative genes by linkage analysis provides substantial information for a molecular diagnosis. To achieve the best results from limited numbers of individuals in a given pedigree, many tightly linked genetic markers with high heterozygosity must be examined, to maximize the chance of analyzing informative markers within a given family.

In the present study, we showed the feasibility of a hierarchical approach for molecular diagnosis of adRP. This approach involved collecting a set of genetic markers that were tightly linked to each of the candidate adRP genes. These markers were then amplified by multiplex polymerase chain reaction (PCR) and subjected to linkage analysis. Based on concordance or discordance of segregation for the markers and disease, candidate genes were selected and mutation screening for the candidate genes was performed.

From the 1Department of Ophthalmology, Fukuoka University School of Medicine, Fukuoka, Japan; the 2Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; and the 3Department of Ophthalmology and Visual Science, Graduate School of Medicine, Chiba University, Chiba, Japan.

Supported by the Japanese Retinitis Pigmentosa Society.

Submitted for publication September 2, 2002; revised October 7, 2002; accepted October 10, 2002.

Disclosure: H. Kondo, None; T. Tabira, None; A. Mizota, None; E. Adachi-Usami, None; K. Oshima, None; K. Hayashi, None.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Hiroaki Kondo, Department of Ophthalmology, Fukuoka University School of Medicine, 7-4-51, Nanakuma, Jonan-ku, Fukuoka, 814-0180, Japan; hkondo@fukuoka-u.ac.jp.
METHODS

Clinical Description of adRP Family

The proband was a 56-year-old man of Japanese origin. He was identified as having night blindness by 10 years of age, and RP had been diagnosed. The best corrected visual acuity was 0.8 in both eyes. The fundus in both eyes showed a degenerative retina with typical peripheral bone spicules (Fig. 1). The visual field was constricted to 10° in both eyes. A dark-adapted electoretinogram (ERG) showed a nonrecordable pattern in both eyes. Two offspring of the proband had received a diagnosis of RP at the ages of 3 and 2 years. The ERG was nonrecordable at the time of diagnosis. The proband had two affected brothers and a healthy sister. One brother, 59 years old, had a time course of disease expression similar to that of the proband, and the visual acuity declined to hand motion in both eyes. The family structure, including other members, is shown in Figure 2. The study followed the tenets of the Declaration of Helsinki. Informed consent was obtained from all subjects involved in the study and the protocol was approved by the review board of Fukuoka University.

Candidate Genes

Of the 13 genes causing adRP, 10 were targeted: RHOD, RDS, CRX, NRL, RP1, FSCN2, PRPF8, PRPF31, HPRP3P, and IMPDH1. The ROM1 gene was not examined, because a recent report indicates only a small contribution of this gene to adRP.21 PIM1K had not yet been cloned at the start of this study. RP17 has yet to be cloned.

Search of Short Tandem Repeats and Primer Design

For each gene, except FSCN2, the human draft sequence data were retrieved from GenBank. The accession numbers of sequence contigs are NT_006025.4 (RHOD), NT_023409.5 (RDS), NT_008101.5 (RP1), NT_019583.5 (NRL), NT_011190.5 (CRX), NT_010692.6 (PRPF8), NT_011148.6 (PRPF31), NT_029226.4 (HPRP3P), and NT_007953.8 (IMPDH1). Short tandem repeats (STRs) were chosen from the gene, basically within 300 kb, if the number of repetitions was more than 13 for dinucleotide repeats and more than 7 for tri- and tetranucleotide repeats. We could not accurately locate the FSCN2 gene on the draft sequence, although the gene locus has been established to be at 17q25. As an alternative way to search for novel markers, six known polymorphic microsatellite markers near the locus: AFMc008we1, AFMa133yg9, GATA65G11, AFM107ve3, AFM210xxs5, and AFM163yg1, encompassing approximately 6 cM, were selected from the Genethon (www.genethon.fr; provided in the public domain by the French Association against Myopathies, Evry, France)22 and The Cooperative Human Linkage Center (CHLC; http://gai.nci.nih.gov/provided in the public domain by the Laboratory for Population Genetics, Division of...
Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health) databases. Four to 12 candidate STRs were selected for each gene, and primer sequences were designed using Primer3 (http://www-genome.wi.mit.edu; provided in the public domain by the Whitehead Institute, Massachusetts Institute of Technology, Cambridge, MA) after masking interspersed repetitive elements with RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker; provided in the public domain by the University of Washington Genome Center, Seattle, WA). One of the 5′-ends of each primer pair was modified for post-PCR fluorescent labeling to contain either a GGT or ATT for the purpose of a 5′-exchange labeling reaction with N,N′-diethyl-2′,7′-dimethyl-6-carboxyfluorescein labeled deoxyguanosine 5′-triphosphate (Roche Applied Science, Foster City, CA) or with 6-carboxyfluorescein labeled deoxyxytidine 5′-triphosphate (R110-dUTP; Applied Biosystems, Foster City, CA). Genomic DNA from seven individuals of the adRP family was used for post-PCR fluorescent labeling, as described previously. An aliquot of each PCR product was examined by agarose gel electrophoresis to confirm specific amplification. The remainder of the amplification products served as substrates for post-PCR fluorescent labeling and fragment length analysis, as described previously. STRs, for which at least one of five individuals was heterozygous, were regarded to be sufficiently polymorphic and were used in further analyses.

Optimization of Multiplex PCR

A set of amplicons for multiplex PCR were created with primers for the STRs of each gene. Initially, primers were simply mixed at equal concentration and a 40-cycle PCR performed, followed by labeling and electrophoresis. Based on the results, primer sets for unsuccessful STR amplifications were eliminated and the PCR was repeated. Concentrations of the primers were then adjusted to obtain peak heights of the same magnitude for all loci in the multiplex analysis.

Fragment Length Analysis

Capillary electrophoresis was performed with a gene sequencer (ABI 310 or 3700; Applied Biosystems) under the conditions suggested by the manufacturer. The dye matrix for the analysis using the 3700 Gene Analyzer was modified and optimized for the dyes used in this study. Data collection and allele identification was performed on computer (GeneScan and Genotyper software; Applied Biosystems).

Genotyping of adRP Family

Seven individuals of the adRP family participated in this study (Fig. 2). DNA was isolated from peripheral blood using a DNA extraction kit (QiAmp; Qiagen, Chatsworth, CA). Genotyping was performed as described earlier, and haplotypes were determined with GeneHunter (http://linkage.rockefeller.edu/soft/gh; provided in the public domain by Rockefeller University, New York, NY). The most likely position of recombinants is shown by horizontal bars or edges of boxes. Note that an allele of marker 4 of RHO seemed not to be amplified and was denoted by 0. An allele of marker 4 of FSCN2 was regarded to be mutated in individual II-6 (*)

---

**FIGURE 2.** Pedigree of the adRP family. Filled symbols: individuals with diagnosed RP; filled boxes: haplotype shared among affected individuals (presumed mutation-carrying haplotype); open box: haplotype identical with the presumed mutation-carrying haplotype but inherited from normal individuals. (A) Genes perfectly cosegregated with disease trait. (B) RP1 showing consistent segregation with the trait, but markers of one individual (II-4) were uninformative and inheritance of a mutated allele to the two siblings was not definitively determined. Hatched box: confounding haplotype. (C) Remainder of the genes, inheritance of which was discordant with the trait, but markers of one individual (II-4) were uninformative and inheritance of a mutated allele to the two siblings was not definitively determined. Hatched box: confounding haplotype.
Mutation Screening
Based on the results of the linkage analysis for each gene locus, remaining candidate genes were analyzed by denaturing high-performance liquid chromatography (DHPLC). 20 The corresponding genomic sequences of the genes were retrieved, and oligonucleotide primers were designed by Primer3, so that each exon was bracketed by genomic sequences of the genes were retrieved, and oligonucleotide primers were designed by Primer3, so that each exon was bracketed by the intronic primers. Before analyzing the patients’ samples, the optimal annealing temperature for PCR was determined empirically. After 30 cycles of PCR in a total volume of 20 μl, the amplification products were applied to DHPLC. Once the DNA fragments for which heterogeneous mobility in DHPLC were segregated with disease trait, direct sequencing was performed with a kit (BigDye Terminator Sequencing; Applied Biosystems) after treatment with shrimp alkaline phosphatase (Amersham). The samples were denatured and analyzed with a DNA sequencer (3700 Gene Analyzer; Applied Biosystems). Mutation in exon 42 of PRPF8 was analyzed with a 217-bp PCR fragment generated with the primers 5’-ATTAGCCAGGCCGGAAC-3’ and 5’-GTGGCGCCTGTTAATGCTCGT-3’, to which the tagged sequences (5’-ATT and 5’-GT) were added for postlabeling purposes. 24

RESULTS
Candidate STRs of RP Genes
Genomic sequences were obtained for nine adRP genes from GenBank. The length of the sequences varied from 286 to 6109 kb (median, 1262). More than half of the STRs identified in the sequences were not suitable for designing primers, because they were juxtaposed with repeated sequences. Except for HPRP3P, the distance of each STR from the RP gene differed from 0 kb (within intron) to 280 kb (average, 110 kb). Only one STR in the contig containing HPRP3P (NT_029226.4, 860 kb) was suitable for the present analysis Therefore, an additional three STRs were picked up from neighboring contig (NT_004811.8) by referring to the Human Genome Browser (http://genome.ucsc.edu/provided in the public domain by UCSC Genome Bioinformatics, University of California at Santa Cruz, Santa Cruz, CA). These were two anonymous STRs in NT_004811.8 and one Genethon marker (AFM135yc3). Four markers located less than 1 cm from HPRP3P were used. Six known STR markers were chosen for FSCN2. The detailed data of the STRs including the primer sequences, precise location, and heterozygosity are available from the first author on request (hkondo@fukouka-u.ac.jp).

Among the 58 STRs of which DNA fragments were successfully amplified, 53 (91%) showed polymorphisms among the five normal individuals. Poor amplification was observed for some of the markers under multiplex PCR conditions. Subsequently, 40 (69%) of 53 polymorphic STRs were used for multiplex amplification. Three to six STRs were used for each subset of multiplex PCR for each of the 10 genes. PCR product size, deduced from draft sequence data, ranged from 107 to 308 bp (average, 176 bp). The 40 markers consisted of 33 dinucleotide, 2 trinucleotide, and 5 tetranucleotide markers, of which 18 markers were labeled by R6G and 22 markers by R110. On the basis of five normal unrelated individuals analyzed, the number of alleles varied from two to six (median, three) and heterozygosity from 0.18 to 0.82 (average, 0.57).

Genotyping of the adRP-Affected Family
Ten PCRs for 10 gene loci were performed for each individual. A computer program (GeneScan; Applied Biosystems) produced clear chromatograms (Fig. 3). Although fluorescent signals for the amplifications differed among STRs, all alleles were successfully genotyped by computer (Genotyper software; Applied Biosystems). Among the 40 STRs genotyped, all but two markers were consistent with Mendelian inheritance. One allele in marker 4 of RHO was not amplified in some individuals. We assumed this was due to the presence of a null allele. One allele of marker 4 of FSCN2 gene was regarded as a mutation in individual II-6, as shown in Figure 2.

Haplotype analysis indicated that the 10 adRP genes examined herein were divided into two categories (Table 1): RHO, RDS, NRL, CRX, PRPF31, HPRP3P, and FSCN2 showed discordant segregation with the disease trait (Fig. 2C), and PRPF8, IMPDH1, and RPI showed perfect cosegregation with the trait (Figs. 2A, 2B). However, as was true of RPI, both haplotypes for individual II-4 were identical. Thus, less information was available with the markers used in this study (Fig. 2B).

Searching for Mutation
From the results just described, we assumed that PRPF8 and IMPDH1 were the candidate genes to be analyzed with high priority. The previous reports suggest that patients with the PRPF8 mutation exhibit early onset of disease symptoms, whereas those with mutation in IMPDH1 have late-onset disease. 27,28 All affected individuals of the family showed an early onset of symptoms. Thus, we started searching for mutation in PRPF8. In particular, we focused on exon 42 of the gene, where previously reported mutations were exclusively located. 15,20

Amplified DNA of exon 42 of PRPF8 showed a heterogeneous mobility in DHPLC in all affected individuals. Fragment size analysis revealed an extra peak 5 bp longer than the normal allele (Fig. 4, left). Sequence analysis revealed a 6-bp deletion and 11-bp insertion of the coding nucleotides at nucleotides 6972 to 6977 (Fig. 4, right). Because of the frame shift of the coding sequence, the last 11 amino acids (codons 2325 to 2335: VYSADREDLYA) were replaced by a longer putative sequence (TLCSLRIGTCMPPFASCFLPRPKPQP-IQTGR) in the mutated allele. The deleted residues contain highly conserved sequence (DRED) across homologous genes in Mus musculus, Drosophila melanogaster, Caenorhabditis elegans, and Arabidopsis thaliana.

DISCUSSION
We showed that a linkage-based exclusion screening, with multiplex linkage analysis with locus-specific microsatellite markers, is an efficient and practical method for narrowing the candidate genes responsible for adRP. Examination of an adRP-affected family by using this strategy indicated that 3 of 10 candidate genes segregated concordantly with the disease. A further search for mutations revealed a novel insertion and deletion in the last exon of a splice-forming factor gene, PRPF8.

With highly accurate human genome sequence data now available in the GenBank public database, it is possible to collect informative markers, such as microsatellites, that are tightly linked to most of the genes. By performance of linkage analysis with these markers, many of the candidate genes responsible for genetically heterogeneous hereditary diseases can be efficiently excluded. The efficiency of exclusion depends on the number of meioses (n) within the examined family. In the case of highly penetrant autosomal dominant diseases such as adRP, incidence of false loci remaining by chance after the exclusion test is given by (1/2)^n, if the phase is known (i.e., if the DNA of grandparents of the affected parent are available for the study). 29 For a nuclear family comprising parents and offspring, the chance is (1/2)^2-1, because the phase is unknown. Thus, for a nuclear family with four children, the probability of failing to exclude nonresponsive genes under such phase-unknown conditions is 1 in 8, if markers associated with the candidate genes are fully informa-
tive. In this study, seven individuals (four plus two meioses with unknown phase) were genotyped, and one (or two) false and one true locus remained, after examination of 10 loci, a number close to that expected.

To facilitate the genotyping, we multiplexed the analysis by PCR. In general, optimization of multiplex PCR is a tedious step because of the interaction of primer sets, which may drastically affect product yield, even if nonspecific amplification and formation of primer dimers are avoided. However, once the optimization is reached, multiplex PCR is a highly efficient method of genotyping many loci. To analyze 10 gene loci, this method requires 10 PCRs and postlabeling per individual, thus the whole genotyping was accomplished by 70 reactions in this study. As an alternative, direct sequencing and/or DHPLC of the common adRP genes in a single proband may be comparable in cost and labor. However, such common genes differ widely in different ethnicities, as between the United States and Japan, and remain unknown in most ethnicities. An advantage of postlabeling in the multiplex genotyping is that sets of markers can be easily revised when the genomic sequence is updated (e.g., FSCN2 and HPRP3P). Toward more efficient use of this method, other candidate genes will be incorporated, such as PIM1K. Although we used 5 individuals to test the polymorphism of microsatellite markers, 50 individuals would give a more accurate representation of the number of different alleles in people and of the heterozygosity fraction. In this study, most markers were located within 500 kb of the candidate genes, so that they are also suitable markers for use in linkage disequilibrium analysis in the future.

McKie et al. identified the gene in the RP13 locus to be PRPF8, a human homologue of the yeast PRP8. PRP8 is expressed ubiquitously and is a component of the spliceosome, an RNA-protein complex consisting of at least 50 proteins and five small nuclear RNAs that are involved in removal of introns of pre-mRNA. The association of the defects in the splicing process with RP is unknown. PRP8 exhibits striking phylogenetic conservation but has no obvious functional motifs, precluding testable predictions for its activity. So far, mutations in RP13 have been shown clustered within a 14-codon stretch in the last exon of PRPF8. The mutations reported previously were all missense and associated with a relatively severe phenotype, suggesting a dominant negative mechanism for disease causation. In this study, we identified a novel mutation of the PRPF8 gene (i.e., a frame-shift mutation that

Table 1. adRP Genes and Haplotype Cosegregation with the adRP Family

<table>
<thead>
<tr>
<th>Haplotype Cosegregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHO</td>
</tr>
<tr>
<td>RDS</td>
</tr>
<tr>
<td>RP1</td>
</tr>
<tr>
<td>NRL</td>
</tr>
<tr>
<td>CRX</td>
</tr>
<tr>
<td>PRPF8</td>
</tr>
<tr>
<td>PRPF31</td>
</tr>
<tr>
<td>HPRP3P</td>
</tr>
<tr>
<td>IMPDH1</td>
</tr>
<tr>
<td>FSCN2</td>
</tr>
</tbody>
</table>
affects C-terminal 11-amino-acid residues). The additional amino acids at the C terminus due to the frame shift could also be responsible for altering the protein in such a way as to cause a dominant negative effect.

Acknowledgments

The authors thank the family members who participated in this study for their cooperation and Hideyuki Hayashi for helpful discussion of the manuscript.

References