Targeted Sequencing of 179 Genes Associated with Hereditary Retinal Dystrophies and 10 Candidate Genes Identifies Novel and Recurrent Mutations in Chinese Patients With Various Retinal Diseases

Supplemental Method

Targeted sequence capture and NGS

6 μg of genomic DNA was randomly fragmented with sizes mainly distributed between 250 and 300 bp. Adapters were ligated to both ends of the resulting fragments. DNA was then amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the RDs189 array for enrichment, and non-hybridized fragments were then washed out. Both non-captured and captured LM-PCR products were subjected to quantitative PCR to estimate the magnitude of enrichment. Each captured library was sequenced by the Illumina Genome Analyser II platform following the manufacturer’s instructions (Illumina, San Diego, CA, USA). Raw image files were processed by the Illumina basecalling Software 1.7 for base calling with default parameters and the sequences of each individual were generated as 90 base paired-end reads.

Bioinformatics analysis

Read alignment and single nucleotide variations (SNVs) calling: For read alignment and SNVs detection, sequencing reads were aligned to the NCBI human reference genome (NCBI build 37.1) using SOAPaligner, a software specifically designed for
Illumina Genome Analyzer technology, as described previously.\(^1\) Reads that were aligned to the designed target regions were collected for SNVs identification and subsequent analysis. The consensus sequence and quality of each allele was calculated by SOAPsnp.\(^2\) Variants were included in subsequent analysis if they met all of the following criteria: (i) quality score ≥20 (Q20); (ii) average copy number at the allelic site ≤2; (iii) distance of two adjacent single nucleotide polymorphisms (SNPs) ≥5 bp; and (iv) sequencing depth ≥4 and ≤500. For detailed information on parameters used in the two algorithms, please refer to [http://soap.genomics.org.cn/](http://soap.genomics.org.cn/).

**Detection of insertion and deletion:** For analyzing insertions and deletions (indel), Burrows–Wheeler Aligner (BWA) was used to map reads onto the reference as described previously.\(^3\) The information of BWA can be found at [http://bio-bwa.sourceforge.net/](http://bio-bwa.sourceforge.net/). Only mapped reads were passed to GATK release 1.5 software (http://www.broadinstitute.org/gsa/wiki/index.php/Home_Page) to identify the breakpoints and to annotate the genotypes of insertions and deletions.

**Sanger sequencing**

Sanger sequencing was used for validation of variations detected by NGS, for segregation analysis of interesting variations when possible, and for prevalence testing of variations in unrelated controls. We have previously described our standard protocol for Sanger sequencing.\(^4\) Primer information for screening mutations of interest is available upon request.
Public Database

For all potential novel pathogenic variations, the PolyPhen-2 (Polymorphism Phenotyping, version 2) Web based service (http://genetics.bwh.harvard.edu/pph2/) and SIFT Human Protein DB (http://sift.bii.a-star.edu.sg/) were used to predict the possible impact of missense variations at protein level.

*Database used for human reference genome are following:*

1. Human reference genome (NCBI build37.1) at UCSC
   
   
   Note: The chromosome coordinates used in our whole analysis corresponded to UCSC Santa Cruz hg19, build 37.1.

2. The CCDS database version was released at April 22, 2011
   

3. RefSeq gene was extracted from refGene.txt file from UCSC
   

4. Ensembl database version was v59
   

*SNP database used for comparison of variations are following:*

1. dbSNP (v132):
http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/snp132.txt.gz

2. exome-sequenced HapMap samples (pilot1, 2, 3):

3. the SNP release of the 1000 Genome Project (20100804 release):
   http://www.1000genomes.org/home

4. YH databases:
   http://yh.genomics.org.cn/

**Mutation database:**

1. Human Gene Mutation Database (HGMD):
   http://www.hgmd.cf.ac.uk/ac/index.php

2. The Genome Database (GDB): http://www.gdb.org

3. Online Mendelian Inheritance in Man (OMIM):


5. KMDB/MutationView:
   http://mutview.dmb.med.keio.ac.jp/MutationView/jsp/index.jsp

**In vitro analysis of RDH5 variant**

*Construction of the targeting vector*

Wide-type *RDH5* cDNA with stop codon removed was synthesized in collaboration
with BGI. *RDH5 p.Arg209* mutation was generated by two-step PCR overlap extension method\(^5,6\) with primers as below. F1:

5’-CTACTAGCTAGCGATGGGTCTTGCCCTTTCTGCT-3’; R1:

5’-TGCTGCACCGGTGGGTAGACTGCTTGCGACGGCTTG-3’; F2:

5’-AGCCTGCTTCTCTAGACCCCTGTGACC-3’; R2:

5’-GGTCACAGGGTTTCTAGAAGCCAGGCTTG-3’.

*RDH5* Arg209 and *RDH5* WT cDNA fragments were cloned into pEGFP-N1 vector (Clontech, Mountain View, CA). The generated constructs were confirmed by direct sequencing.

**Cell culture and transient transfection**

Human embryonic kidney 293 (HEK293) cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 g/ml streptomycin at 37°C and 5% CO\(_2\). At 80–90% confluency, cells were transiently transfected using Lipofectamine\(^\text{TM}\) 2000 Transfection reagent (Invitrogen, Carlsbad, CA), with 4µg of pEGFP-*RDH5* Arg209\(^*\) or pEGFP-*RDH5* WT, or empty vectors. Transfected cells were visualized by Green fluorescent protein (GFP) fluorescence.

**RT-PCR analysis of RDH5 expression**

At 48 hours after transfection, total RNA was isolated with Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions, and further treated with DNase I (Takara, Dalian, China). Reverse transcription-PCR (RT-PCR) was performed as previously described,\(^7\) with primers specific for *RDH5* transcript (F:
5’-TGGGAGGTCTATAAGCAGAG-3’, R:
5’-CCATTGTGTTTCACATTTCAGCA-3’) (Figure 5). To determine the transfection efficiency, genomic DNA was also extracted from each type of transfected cells. Semiquantitative PCR amplification (25 cycles, 10 sec at 98°C, 15 sec at 60°C and 2 min at 72°C) of DNA template was performed with primers specific for the RDH5-GFP fusion transcript (F: 5’-TGGGAGGTCTATAAGCAGAG-3’, R: 5’-CTACAAATGTGGTTATGGCTGATT-3’).

*Immunoblot analysis*

At 72 hours after transfection, cells were harvested and proteins were extracted from cell lysates. Equal amount of each protein samples were resolved by 10% SDS-PAGE gel. Antibodies against RDH5 (1:1000), GFP (1:5000) and GAPDH (1:5000) were purchased from Abcam (Abcam, Hong Kong).
References:


