Published in final edited form as:

Adv Exp Med Biol. 2010 ; 664: 325–331. doi:10.1007/978-1-4419-1399-9_37.

Targeted High-Throughput DNA Sequencing for Gene Discovery in Retinitis Pigmentosa

Stephen P. Daiger, Lori S. Sullivan, Sara J. Bowne, David G. Birch, John R. Heckenlively, Eric A. Pierce, and George M. Weinstock

Abstract

The causes of retinitis pigmentosa (RP) are highly heterogeneous, with mutations in more than 60 genes known to cause syndromic and non-syndromic forms of disease. The prevalence of detectable mutations in known genes ranges from 25 to 85%, depending on mode of inheritance. For example, the likelihood of detecting a disease-causing mutation in known genes in patients with autosomal dominant RP (adRP) is 60% in Americans and less in other populations. Thus many RP genes are still unknown or mutations lie outside of commonly tested regions. Furthermore, current screening strategies can be costly and time-consuming.

We are developing targeted high-throughput DNA sequencing to address these problems. In this approach, a microarray with oligonucleotides targeted to hundreds of genes is used to capture sheared human DNA, and the sequence of the eluted DNA is determined by ultra-high-throughput sequencing using next-generation DNA sequencing technology. The first capture array we have designed contains 62 full-length retinal disease genes, including introns and promoter regions, and an additional 531 genes limited to exons and flanking sequences. The full-length genes include all genes known to cause at least 1% of RP or other inherited retinal diseases. All of the genes listed in the RetNet database are included on the capture array as well as many additional retinal-expressed genes. After validation studies, the first DNA's tested will be from 89 unrelated adRP families in which the prevalent RP genes have been excluded. This approach should identify new RP genes and will substantially reduce the cost per patient.

37.1 Introduction

The genetic causes of inherited retinal diseases, even a "simple" category such as autosomal dominant retinitis pigmentosa (adRP), are extremely heterogeneous. More than 190 genes causing inherited retinal diseases have been identified (Fig. 37.1), including at least 40 causing non-syndromic retinitis pigmentosa and 20 causing syndromic forms of RP (Daiger et al. 2007; RetNet 2009). In addition to many disease-causing genes, there are often many different mutations at each locus, and different mutations within the same gene may cause strikingly different diseases. Further, in spite of the large number of genes identified to date, the fraction of patients in which a mutation can be found by screening the known genes is often low. For example, screening known genes in adRP families leads to identification of a disease-causing mutation in 60% of cases among Americans of European origin and less frequently among other populations (Fig. 37.2). Thus there are many retinal disease genes that have not been identified yet.

Next generation sequencing techniques, that is, novel gene selection and targeting methods followed by massively-parallel, ultra-high-throughput sequencing, offer a rapid, efficient way to find disease-causing mutations in affected individuals and to discover new disease genes (Albert et al. 2007). We are applying these methods to finding genes and mutations causing adRP, focusing on a cohort of 89 families in which conventional testing failed to detect mutations in known genes (Sullivan et al. 2006; Sullivan et al. 2006a; Gire et al. 2007; Bowne et al. 2008). That is, these families have mutations in novel adRP genes or mutations in known genes that are not readily detectable, for example, mutations outside of coding regions.

Our approach is to address these possibilities by targeting a large number of known and candidate retinal disease genes using oligonucleotide capture arrays followed by ultra-high-throughput sequencing. In addition to the many candidates for gene discovery, the capture arrays include non-coding sequences of known retinal disease genes to detect subtle mutations. We refer to this approach as the VisionCHIP. 'VisionCHIP' stands for Comprehensive High-Throughput Interrogation of Patient DNAs for Vision Research.

The first disease targeted for study is adRP – because of the availability of families enriched in novel genes and mutations, and because many of the families are large enough to test segregation of potentially pathogenic mutations, a major problem in assessing rare variants. However, the VisionCHIP approach, once optimized and validated, will be equally applicable to other forms of inherited retinal disease.

37.2 Methods

37.2.1 Selection of Families

In earlier and continuing research, we have ascertained and acquired DNA samples from over 500 families with a diagnosis of adRP (Sullivan et al. 2006). Among these, we have selected families with at least (i) three generations of inheritance and multiple affected females or (ii) two affected generations, three or more affected individuals and male-to-male transmission. That is, these families are more likely to have dominant RP and less likely to have an X-linked mode of inheritance. This is our adRP cohort; at present, there are 228 families in the cohort, approximately 85% white, 5% Hispanic, 5% African American, and 5% Asian and other.

The 228 families in the adRP cohort have been screened for mutations by a number of methods: sequencing of known genes (Sullivan et al. 2006), deletion testing using multiplex ligation-dependent probe amplification (MLPA) (Sullivan et al. 2006a), linkage mapping (Sullivan et al. 2005) and candidate gene screening (Gire et al. 2007; Bowne 2008). To date we have found mutations in 61% of these families (Fig. 37.2 and unpublished), leaving 89 for gene discovery. The additional adRP patients who are not part of the cohort are available for further screening of likely candidate genes.

37.2.2 VisionCHIP Gene Selection

Version 1 of the VisionCHIP contains 593 genes divided into three categories in terms of sequence coverage:

1. Genes Less Than 100 Kb In Length, Known To Cause Some Form Of Retinal Degeneration, Which Will Be Sequenced Completely (51 Genes);
2. Genes Larger Than 100 Kb In Length, Known To Cause Some Form Of Retinal Degeneration, Which Will Have All Exons And Some Non-Coding Regions Sequenced (11 Genes); And

3. Genes That Are Potential Candidates For Retinal Degeneration Which Will Have Exons And Exon-Flanking Regions Sequenced (531 Genes).

Genes in categories 1 and 2 were derived from the RetNet database of retinal disease genes (RetNet 2009). Genes that are known to cause 1% or more of cases of retinitis pigmentosa, juvenile macular degeneration, or cognate diseases were selected for full-length sequencing because these genes are most likely to have disease-causing mutations, and some mutations may fall outside of coding regions (Daiger et al. 2007).

Genes in category 3 came from multiple sources, including the EyeSAGE database (Bowes Rickman et al. 2006), and the human homologs of genes coding for proteins found in mouse photoreceptor outer segments and axonemes (Liu et al. 2007). Additional candidates were chosen from the retinal literature, while others were found in public databases such as NEIBank (2008), UniGene (2008), Entrez (2008), the Human Protein Reference Database (2008), and BioGRID (2008). Characteristics of chosen genes include high levels of retina/photoreceptor/eye/cilia expression; interaction with known disease genes; sequence similarity to known retinal disease genes; identification in screens of retinal gene expression; similarity in expression patterns to known retinal disease genes; candidate genes proposed by other investigators; and genes previously tested in our laboratory as potential candidates.

Figure 37.3 shows the chromosomal distribution of the first set of genes chosen for the VisionCHIP.

37.2.3 VisionCHIP Validation

To optimize and validate the VisionCHIP, we are focusing on controls with known adRP mutations, including deletions, and on 21 families from the adRP cohort without known mutations. The 21 families each have multiple affected members immediately available for segregation testing. In addition, three of the largest families are being tested for genome-wide linkage using Affymetrix 6.0 SNP Arrays. The linkage testing will provide genotypes for independent validation of SNPs within VisionCHIP genes, and may implicate linkage regions containing targeted retinal genes.

The current iteration of the VisionCHIP is being fabricated by NimbleGen Inc. (Roch). An alternative capture method is under development at the Genome Sequencing Center, Washington University (WU-GSC), St. Louis. Patient DNAs are subjected to whole-genome amplification, and then sheared, ligated with universal primers and individual 'bar codes', pooled and captured. The eluted, targeted DNA is then amplified and sequenced using 454 FLX (Roche), 454 Titanium (Roche) and/or Solexa (Illumina) ultra-high-throughput, massively parallel sequencers. Sequencing and sequence assembly are underway at WU-GSC. We anticipate 5–10 Mb of diploid sequence, 30–50-fold depth, for nearly 600 retinal genes, from each of the 21 families. In practice, we are actually testing pairs of affected individuals from each family (as far apart in the pedigree as possible) to generate preliminary segregation information for each variant observed.

37.2.4 Evaluating Potentially Pathogenic Variants

Because of the extensive sequencing of retinal genes, including introns and promoter regions, we anticipate that a significant fraction of patients will be found to have novel, rare variants in known RP genes and rare variants in many candidate genes. We are focusing on the RP genes first. Bioinformatics analysis of amino acid substitutions involves application of PolyPhen and related programs (Grantham 1974; Ng and Henikoff 2003; Stone 2007). Intronic sequences will be examined for possible splice-altering mutations using a combination of NNSPLICE and ASSP (alternative splice site predictor) (Wang and Marin

2006). Promoter regions will be determined and analyzed using programs such as PromoterInspector and Dragon Promoter Finder (Bajic et al. 2002; Scherf et al. 2001). Additional computational methods for ranking possible pathogenicity are in Sullivan et al. (2006). Investigators at the WU-GSC have successfully applied a suite of computational tools to identify pathogenic somatic mutations in adenocarcinoma (Ding et al. 2008).

In order to detect copy number variants (CNVs), especially deletions, we will examine genotypes of SNPs within each gene, looking for extended regions of homozygosity. We also plan to work with WU-GSC to look for gene regions that appear to be over or under represented and to determine if this correlates with CNVs. Our existing panel of adRP patients with large PRPF31 deletions will be used as controls (Sullivan et al. 2006a).

37.3 Conclusion

The VisionCHIP approach to finding retinal disease genes, based on targeted capture and ultra-high-throughput sequencing, is only one step towards whole-genome sequencing to identify mutations causing single-gene Mendelian disorders. Whole-genome sequencing will become widely available in as little as 5 years, possibly based on single-molecule techniques. The problem lies not with the sequencing technology but with analyzing and understanding the resulting genotypes. Humans are heterozygous for a nucleotide substitution roughly every 1,000 bp; of these, roughly 1 in 20 are rare, non-polymorphic variants. With the conservative estimate that 1 in 10 of *these* are potentially pathogenic based on computational analysis, each individual will be heterozygous for a potential disease-causing variant every 200 kb. Therefore, we anticipate detecting dozens of possible mutations per person with the first iteration of the VisionCHIP, and thousands when whole-genome sequences become routinely available. This estimate does not include indels, copy number variants, variable repeats or other DNA variants.

Determining pathogenicity of rare variants will be a major challenge to medical genetics for the foreseeable future. We believe that autosomal dominant retinitis pigmentosa offers a model system for addressing this problem. First, even though many adRP genes are not yet known, many disease-causing genes and mutations, and disease pathways, have been identified already. Also, retinal biology is a highly developed science. Thus there is a strong scientific background against which to judge novel genes and variants. Second, many multi-generation adRP families are available for segregation analysis – perhaps the most powerful means to assess pathogenicity. Finally, there are many functional assays for retinal gene mutations including in vitro systems, single-cell models, animal models and powerful imaging techniques for localizing and characterizing retinal proteins. Taken together, these approaches are likely to reveal several or many new adRP genes and many novel mutations.

Acknowledgments

Supported by grants from the Foundation Fighting Blindness, The Gustavus and Louise Pfeiffer Research Foundation, the Herman Eye Fund, and NIH grants EY007142 and EY005235.

References

- Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;4:903–905. [PubMed: 17934467]
- Bajic VB, Seah SH, Chong A, et al. Dragon promoter finder: recognition of vertebrate RNA polymerase II promoters. *Bioinformatics* 2002;18:198–199. [PubMed: 11836231]
- BioGRI, D. A general repository for interaction datasets. 2008 [December 1]. <http://www.thebiogrid.org/>.

- Bowes Rickman C, Ebright JN, Zavodni ZJ, et al. Defining the human macula transcriptome and candidate retinal disease genes using EyeSAGE. *Invest Ophthalmol Vis Sci* 2006;47:2305–2316. [PubMed: 16723438]
- Bowne SJ, Sullivan LS, Gire AI, et al. Mutations in the TOPORS gene cause 1% of autosomal dominant retinitis pigmentosa (adRP). *Mol Vis* 2008;14:922–927. [PubMed: 18509552]
- Daiger SP, Bowne SJ, Sullivan LS. Perspective on genes and mutations causing retinitis pigmentosa. *Arch Ophthalmol* 2007;125:151–158. [PubMed: 17296890]
- Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–1075. [PubMed: 18948947]
- Entrez, G. NCBI Entrez Gene database. 2008 [December 1]. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>.
- Gire AI, Sullivan LS, Bowne SJ, et al. The Gly56Arg mutation in NR2E3 accounts for 1–2% of autosomal dominant retinitis pigmentosa. *Mol Vis* 2007;13:1970–1975. [PubMed: 17982421]
- Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974;185:862–864. [PubMed: 4843792]
- Human Protein Reference Database. HPRD. 2008 [December 1]. <http://www.hprd.org/>.
- Liu Q, Tan G, Levenkova N, et al. The proteome of the mouse photoreceptor sensory cilium complex. *Mol Cell Proteomics* 2007;6:1299–1317. [PubMed: 17494944]
- NEIBank. NEI database of eye tissue ESTs. 2008 [December 1]. <http://neibank.nei.nih.gov/>.
- Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–3814. [PubMed: 12824425]
- RetNet. The Retinal Information Network. 2009 [December 1]. <http://www.sph.uth.tmc.edu/RetNet/>. Stephen P. Daiger, PhD, Administrator, The Univ. of Texas Health Science Center at Houston.
- Scherf M, Klingenhoff A, Frech K, et al. First pass annotation of promoters on human chromosome 22. *Genome Res* 2001;11:333–340. [PubMed: 11230158]
- Stone EM. Leber congenital amaurosis – a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. *Am J Ophthalmol* 2007;144:791–811. [PubMed: 17964524]
- Sullivan LS, Bowne SJ, Birch DG, et al. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa (adRP): a screen of known genes in 200 families. *Invest Ophthalmol Vis Sci* 2006;47:3052–3064. [PubMed: 16799052]
- Sullivan LS, Bowne SJ, Seaman CR, et al. Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 2006a;47:4579–4588. [PubMed: 17003455]
- Sullivan LS, Bowne SJ, Shankar SP, et al. Linkage mapping in families with autosomal dominant retinitis pigmentosa (adRP). *Invest Ophthalmol Vis Sci* 2005;46 E-Abstract 2293.
- UniGene. NCBI UniGene database. 2008 [December 1]. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>.
- Wang M, Marin A. Characterization and prediction of alternative splice sites. *Gene* 2006;366:219–227. [PubMed: 16226402]

Fig. 37.1.

Graph of mapped and identified retinal disease genes from 1980, the beginning of the modern era of gene discovery, through December 2008 (RetNet 2009)

Fig. 37.2.

Fraction of mutations detected in known adRP genes in a cohort of 228 adRP families (Sullivan et al. 2006; Sullivan et al. 2006a; Gire et al. 2007; Bowne et al. 2008; and unpublished)

Fig. 37.3.

Map location of 593 genes chosen for inclusion on the first iteration of the VisionCHIP