



HAL
open science

Molecular diagnosis for heterogeneous genetic diseases with high-throughput DNA sequencing applied to retinitis pigmentosa

David Arthur Simpson, Graeme Richard Clark, Sharon Alexander, Giuliana Silvestri, Colin Eric Willoughby

► To cite this version:

David Arthur Simpson, Graeme Richard Clark, Sharon Alexander, Giuliana Silvestri, Colin Eric Willoughby. Molecular diagnosis for heterogeneous genetic diseases with high-throughput DNA sequencing applied to retinitis pigmentosa. *Journal of Medical Genetics*, 2010, 48 (3), pp.145. 10.1136/jmg.2010.083568 . hal-00591224

HAL Id: hal-00591224

<https://hal.science/hal-00591224>

Submitted on 8 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITLE PAGE

Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa

David A. Simpson, Graeme R Clark, Sharon Alexander, Giuliana Silvestri, Colin E. Willoughby

Corresponding author: David Arthur Simpson

Queen's University Belfast, Centre for Vision & Vascular Science, Ophthalmic Research Centre, Institute of Clinical Science A, Royal Victoria Hospital, Belfast BT12 6BA, N. Ireland, UK

Email: David.Simpson@qub.ac.uk

Tel: +44 28 90632719 Fax: +44 28 90632699

Co-authors: Graeme Richard Clark, Sharon Alexander, Giuliana Silvestri, Colin Eric Willoughby

Centre for Vision and Vascular Science, Queen's University Belfast, Belfast, Northern Ireland, BT12 6BA, UK

KEYWORDS

Diagnostics test, clinical genetics, Genetic screening/counselling, Molecular genetics, Ophthalmology

Word count: 2802

ABSTRACT

Background

The genetic heterogeneity of many Mendelian disorders, such as retinitis pigmentosa which results from mutations in over 40 genes, is a major obstacle to obtaining a molecular diagnosis in clinical practice. Targeted high-throughput DNA sequencing offers a potential solution and was used to develop a molecular diagnostic screen for retinitis pigmentosa patients.

Methods

A custom sequence capture array was designed to target the coding regions of all known retinitis pigmentosa genes and used to enrich these sequences from DNA samples of five patients. Enriched DNA was subjected to high-throughput sequencing singly or in pools and sequence variants were identified by alignment of up to 10 million reads per sample to the normal reference sequence. Potential pathogenicity was assessed by functional predictions and frequency in controls.

Results and conclusions

Known homozygous *PDE6B* and compound heterozygous *CRB1* mutations were detected in two patients. A novel homozygous missense mutation (c.2957A>T; p.N986I) in the cyclic nucleotide gated channel beta 1 (*CNGB1*) gene predicted to have a deleterious effect and absent in 720 control chromosomes was detected in one case in which conventional genetic screening had failed to detect mutations. The detection of known and novel retinitis pigmentosa mutations in this study establishes high-throughput DNA sequencing with DNA pooling as an effective diagnostic tool for heterogeneous genetic diseases.

INTRODUCTION

The completion of the Human Genome Project in 2003 was heralded as the dawn of an era of genomic medicine,¹ in which information from genomes would guide clinical decision making and deliver personalized medicine.² It was anticipated that accelerated detection of disease-related mutations would improve genetic diagnosis and prognosis.³ However, delivery of personalized genomic medicine requires not only access to the complete human genome, but availability of appropriate genetic tests for individual patients. The genetic heterogeneity of many Mendelian disorders is a major obstacle to obtaining molecular diagnoses in clinical practice.⁴ For example, Retinitis Pigmentosa⁵ (RP[MIM #268000]), the commonest inherited retinal degeneration (prevalence 1:4000), is caused by mutations in over 40 genes for the non-syndromic form of the disease alone (Retnet: Retinal Information Network, <http://www.sph.uth.tmc.edu/retnet/>). Molecular genetic testing is important for clinical care,⁶ enabling assignment of risk, genetic counseling and prognosis, and will be essential for enrolling patients in the future gene therapy trials likely to stem from the promising current human trials of *RPE65* (MIM +180069) therapies for Leber congenital amaurosis.^{7,8}

Apart from the genetic heterogeneity of RP, there are a number of obstacles that currently limit molecular diagnosis and therefore hinder the potential for personalized genomic medicine to guide clinical decision making. The lack of clearly defined genotype-phenotype correlations makes it difficult to direct testing to specific candidate genes. Approximately 50% of all RP patients are the only known affected family member and have no evidence of consanguinity; these patients are classified as isolated, sporadic or simplex cases⁹ which excludes diagnostic evaluation based on inheritance pattern. Traditional genetic screening for RP is laborious, although technological advances have had some impact.¹⁰⁻¹² The

development of massively parallel or ‘next-generation’ sequencing techniques which generate millions of DNA sequence reads in parallel during a single experimental run offers a potential solution.^{4,13} However, to date, applications of this high-throughput DNA sequencing approach as a molecular diagnostic tool have been limited due to the costs and perceived technical and data handling challenges.^{14,15} The aim of this work is to demonstrate that high-throughput DNA sequencing is now sufficiently established to warrant introduction into clinical practice. Herein we report the use of this technique to detect known and novel mutations in RP patients, demonstrating the clinical utility of this new technology.

METHODS

Participants

All applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during this research. The study was approved by the Northern Ireland Research Ethics Committee and all participants gave informed consent. Patients were selected from a previously reported cohort¹² and all underwent a comprehensive ocular evaluation including Snellen visual acuity testing, visual fields (Humphrey 24-2; Carl Zeiss Meditec, Inc., Dublin, CA, USA), fundal examination, and electro-retinography to establish the diagnosis of RP based on the presence of the typical fundal features (figure 1), visual field constriction and an attenuated or abolished electro-retinogram. DNA was extracted from whole blood using standard protocols (Wizard DNA Purification Kit; Promega, Southampton, UK). 360 DNA samples from the general Northern Ireland population (unscreened for ocular disease) were used as controls.

Sequence capture and next generation sequencing

A custom sequence capture array (Roche NimbleGen, Madison, WI, USA) was designed to target all exons and 100bp of flanking sequence from all genes in which mutations were known to cause RP (Retnet: Retinal Information Network) and selected genes associated with Leber congenital amaurosis, a disease which has phenotypic overlap with RP (table 1).

Table 1. Genes implicated in RP (or Leber congenital amaurosis which shows phenotypic overlap) arranged by inheritance pattern. ^aGenes discovered after the design of the sequence capture array (November 2008) and therefore not included.

Autosomal dominant		Autosomal recessive		X-linked
<i>CA4</i>	<i>RP9</i>	<i>ABCA4</i>	<i>NRL</i>	<i>RPGR</i>
<i>CRX</i>	<i>SEMA4A</i>	<i>AIPL1</i>	<i>PDE6A</i>	<i>RP2</i>
<i>FSCN2</i>	<i>SNRNP200^a</i>	<i>C2ORF71^a</i>	<i>PDE6B</i>	
<i>GUCA1B</i>	<i>TOPORS</i>	<i>CDHR1</i>	<i>PRCD</i>	
<i>IMPDH1</i>		<i>CEP290</i>	<i>PROM1</i>	
<i>KLHL7^a</i>		<i>CERKL</i>	<i>RBP3^a</i>	
<i>NR2E3</i>		<i>CNGA1</i>	<i>RD3</i>	
<i>PRPF3</i>		<i>CNGB1</i>	<i>RGR</i>	
<i>PRPF8</i>		<i>CRB1</i>	<i>RLBP1</i>	
<i>PRPF31</i>		<i>EYS^a</i>	<i>RPE65</i>	
<i>PRPH2</i>		<i>GUCY2D</i>	<i>RPGRIP1</i>	
<i>RDH12</i>		<i>IDH3B</i>	<i>SAG</i>	
<i>RHO</i>		<i>LCA5</i>	<i>SPATA7^a</i>	
<i>ROM1</i>		<i>LRAT</i>	<i>TULP1</i>	
<i>RPI</i>		<i>MERTK</i>	<i>USH2A</i>	

All reference sequences were based upon the NCBI36/hg18 assembly of the human genome. The array comprised 385000 unique probes selected using the Sequence Search and Alignment by Hashing Algorithm¹⁶ to capture a total of 359kb of genomic sequence comprising 681 exons from 45 genes. Patient DNA samples were enriched for the targeted sequences using the manufacturer's protocols. Briefly, 21µg aliquots were fragmented and hybridized to the array, non-target sequences washed off and the enriched fragment DNA pool subsequently eluted and amplified by ligation-mediated PCR. Approximately 5µg of amplified enriched DNA was used as input for massively parallel sequencing on a Genome Analyzer II (Illumina, San Diego, CA, USA) either with a single or four pooled, bar-coded samples per flow cell to generate single end reads of 40bp or 32bp after removal of tags (GATC Biotech, Konstanz, Germany). Reference gene sequences were annotated with known single nucleotide polymorphisms (SNPs) from the NCBI dbSNP database build 130 and RP mutations from The Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/>) or reported in the literature. Sequencing reads were aligned to the reference sequences using Genomic Workbench software (CLC bio, Nottingham, England) with default settings. Sequence variants present in greater than 30% of reads at positions covered by at least five reads were classified as novel or known non-coding, synonymous or non-synonymous changes by comparison with the annotated reference sequences. These thresholds were chosen to maximize the detection of real variants whilst minimising artefactual calls and were informed by observation of the detection rate of known SNPs. Identified sequence variants were annotated according to the guidelines published by the Human Genome Variation Society.

Determination of pathogenicity

The potential pathogenicity of sequence variants identified by high-throughput DNA sequencing was assessed as follows. Firstly, known SNPs were excluded as described above. The frequency of non-synonymous sequence variants in 360 population controls (720 chromosomes) was then determined by either Sequenom mass array genotyping (Sequenom, San Diego, CA, USA) or high resolution melt curve analysis on a LightCycler 480 (Roche Diagnostics, Basel, Switzerland) using Precision HRM MasterMix (Primerdesign, Southampton, UK), according to the manufacturers' protocols. The functional effects of non-synonymous sequence variants that were absent from or present in less than 1% of the control population were predicted using a range of freely-available computational tools. The conservation of the affected amino acid across species was analyzed using the ClustalW multiple sequence alignment program (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Deleterious structural effects of amino acid substitutions on protein function were assessed using the PolyPhen (*Polymorphism Phenotyping*; <http://genetics.bwh.harvard.edu/pph/>), PMut (<http://mmb2.pcb.ub.es:8080/PMut/>) and SIFT (<http://sift.jcvi.org/>) algorithms. Synonymous and intronic sequence variants were assessed for potential deleterious effects upon messenger RNA splicing using the Human Splicing Finder version 2.4 tool (<http://www.umd.be/HSF/>). Sequence variants of interest identified by high-throughput DNA sequencing were verified in replicate PCR amplicons using Sanger sequencing. When other family members were available, the segregation of sequence variants with the disease was assessed.

RESULTS

We performed array-based enrichment of 45 RP genes and subsequent high-throughput DNA sequencing and report the detection of known RP mutations and a novel recessive mutation in *CNGBI* (MIM*600724; c.2957A>T; p.N986I) in a patient with simplex RP. Two patients with an established genetic diagnosis¹² but different patterns of molecular involvement, one homozygote (patient RP148: *PDE6B*; MIM +180072; c.1685G>A; p.G562D) and one compound heterozygote (patient RP179: *CRBI* MIM +604210; c.2129A>T; p.E710V and c.2548G>A; p.G850S), were included to demonstrate the ability of our approach to detect known mutations. Three patients without a genetic diagnosis were selected because they are the only affected member in their family (simplex RP) and conventional genotyping and resequencing arrays had failed to detect mutations¹². Such cases present the greatest challenge and these patients were chosen deliberately to assess the ability of targeted sequence capture and high-throughput DNA sequencing to detect novel mutations in the known RP genes.

A custom sequence capture array was designed to capture all exons and 100bp of flanking sequence from a total of 45 genes in which mutations have been reported to cause RP or Leber congenital amaurosis (table 1). When the captured DNA was subjected to high-throughput DNA sequencing approximately 35% of the reads aligned uniquely to the reference sequences, indicative of a greater than 3000 fold enrichment of the targeted sequences. One sample was sequenced in a single flow cell and resulted in high fold coverage of almost the entire targeted region, with a mean coverage of 486 reads and > 20 fold coverage over more than 99% of the target region (figure 2). The four remaining samples

were labeled with 4bp tags and pooled prior to sequencing. Although the coverage was reduced to an average of 98 reads, approximately 94% and 88% of the target region was covered by at least 10 or 20 reads respectively (figure 2), demonstrating that sample pooling is a feasible approach to reduce sequencing costs.

Applying a requirement for at least five reads and a minimum variant frequency of 30%, 582 single nucleotide variants were detected, of which 150 were novel variants and 432 were previously reported SNPs. Of the novel variants, 113 occurred in non-coding and 37 in coding sequences (13 of which were non-synonymous) and none affected canonical splice sites. The non-synonymous sequence variants detected included the known homozygous and compound heterozygous mutations, and were all confirmed by Sanger sequencing (table 2). The known *PDE6B* homozygous mutation (c.1685G>A; p.G562D) in patient RP148 was detected in all 134 reads which spanned this position (figure 3). For the known *CRB1* heterozygous variants 76 of 135 reads (c.2129A>T; p.E710V) and 78 of 130 reads (c.2548G>A; p.G850S) carried the mutated sequence in patient RP179 (figure 3).

Table 2. Classification of non-synonymous sequence variants detected by massively parallel DNA sequencing. Thirteen non-synonymous variants were detected by massively parallel DNA sequencing including 4 mutations (bold text): one known homozygous mutation in *PDE6B* (patient RP148; p.G562D), 2 compound heterozygous mutations in *CRBI* (patient RP179; p.E710V and p.G850S) and a novel homozygous mutation in *CNGBI* (RP167; p.N986I). ^a The frequency of non-synonymous sequence variants was assessed in 360 population controls (720 chromosomes). ^b Two variants (*ROM1*: p.R229H; *RPI*: p.L1417P) were classified as rare polymorphisms as they were seen in <1% of controls but did not segregate. ^c Variants present in at least 1% of the control population were classified as SNPs. ^d PolyPhen classification: Prob D (probably damaging: high confidence change affects protein function or structure); Poss D (possibly damaging: may affect protein structure or function); and B (benign: lacks phenotypic effect). ^e PMut classification: N (neutral) and P (pathogenic). ^f SIFT classification: T (tolerated) and D (damaging).

Patient	Gene	Exon	Nucleotide Change	Amino Acid Change	Computational Pathogenicity Prediction			Frequency (%) Controls ^a		Variant Classification
					PolyPhen ^d	Pmut ^e	SIFT ^f	WT	Variant	
RP179	<i>CRB1</i>	7	c.2129A>T	p.E710V	Prob D	N	T	99.2	0.8	Known compound heterozygous mutation(12)
	<i>CRB1</i>	7	c.2548G>A	p.G850S	Prob D	N	T	100	0	Known compound heterozygous mutation(17)
RP148	<i>PDE6B</i>	13	c.1685G>A	p.G562D	Poss D	P	D	100	0	Known homozygous mutation(12)
	<i>FSCN2</i>	1	c.538C>T	p.R180W	Prob D	P	T	99.9	0.1	Unclassified variant
RP167	<i>CNGB1</i>	29	c.2957A>T	p.N986I	Prob D	P	D	100	0	Novel homozygous mutation
RP75	<i>ROM1</i>	2	c.686G>A	p.R229H	B	N	T	100	0	Known polymorphism(18) ^b
RP141	<i>ABCA4</i>	23	c.3352C>G	p.H1118D	Prob D	P	D	100	0	Unclassified variant
	<i>CA4</i>	7	c.700G>A	p.V234I	B	N	T	95.4	4.6	SNP ^c
	<i>CNGA1</i>	2	c.95G>A	p.R32Q	B	N	T	98.1	1.9	SNP ^c
	<i>PDE6A</i>	16	c.1963C>T	p.H655Y	B	N	T	97.4	2.6	SNP ^c
	<i>RPI</i>	4	c.4250T>C	p.L1417P	Poss D	P	T	100	0	Rare polymorphism ^b (no segregation) (19)
	<i>RPI</i>	4	c.1118C>T	p.T373I	Poss D	P	D	99	1	SNP ^c (no segregation) (19)
	<i>USH2A</i>	15	c.3123C>A	p.H1041Q	Prob D	N	T	99.7	0.3	Unclassified variant

Two previously reported rare missense variants were detected in *RPI* (MIM *603937), c.4250T>C; p.L1417P¹⁷ in patient RP141 and c.1118C>T; p.T373I¹⁷ in patient RP179. These were absent from or present at a frequency of 1% respectively in 360 control samples (720 chromosomes), but neither segregated with disease and they were therefore considered to be non-pathogenic. Of the remaining eight non-synonymous variants five were found to occur in <1% of controls (table 2). Three of these were novel heterozygous changes (*ABCA4*; MIM *601691; c.3352C>G; p.H1118D; *FSCN2*; MIM *607643; c.538C>T; p.R180W; *USH2A*; MIM +608400; c.3123C>A; p.H1041Q) and because no second potentially pathogenic variant could be found in the respective genes, they were considered unclassified variants. The non-synonymous variant in *ROM1* (MIM *180721; c.686G>A; p.R229H) was found in <1% in controls but was a previously reported polymorphism¹⁸ and no mutation was found in the peripherin-RDS gene (MIM *179605).

One novel homozygous missense variant was detected in a simplex patient (RP167; figure 1), in the cyclic nucleotide gated channel beta 1 gene (*CNGBI*) and was absent in controls (table 2, figure 4). The c.2957A>T sequence variant in *CNGBI* exon 29 results in the non-conservative substitution (p.N986I) of the amidic amino acid asparagine (AAC: Asn; N) with the aliphatic isoleucine (ATC: Ile; I). Computational analyses (PolyPhen, PMut and SIFT) predicted pathogenicity (table 2) and evolutionary conservation analysis (ClustalW) showed the N986 position to be invariant across species with *CNGBI* orthologs (figure 4). When a recessive RP mutation was detected by high-throughput DNA sequencing the affected patient was counseled about the diagnosis and risk of disease transmission. During this interview the patient revealed

that his parents were second cousins, establishing consanguinity in the pedigree. Available family members were screened using Sanger sequencing to assess segregation and the heterozygous state of these family members indicates segregation of the mutation with disease (figure 4).

DISCUSSION

The potential of targeted sequence capture and high-throughput DNA sequencing to improve genetic diagnosis, counseling and deliver personalized medicine is widely acknowledged. However, applications of this approach as a molecular diagnostic tool have been limited and most commentators report that a number of technical and economic issues must be solved before this technique can be incorporated into routine clinical care.^{4,13,15,20} The detection of known and novel RP mutations in this study now establishes DNA pooling and high-throughput DNA sequencing as an effective diagnostic tool for heterogeneous genetic diseases.

Over 40 genes have been identified in RP and there are insufficient phenotypic differences between patients (figure 1) to indicate which gene is likely mutated and guide molecular diagnostics. Inheritance pattern can direct genetic testing, particularly for X-linked RP in which mutations in *RPGR* (MIM *312610) and *RP2* (MIM *300757) predominate, but the majority of patients have no family history.⁹ Current molecular testing strategies focus on small sets of genes and apply conventional laboratory techniques based on the relative frequency of mutations and cost-effectiveness of testing. Commercially available genotyping arrays (Asper Ophthalmics, Tartu, Estonia) will only screen for known mutations based on inheritance pattern.¹⁰) Our results

demonstrate that high-throughput DNA sequencing can detect previously reported mutations in known RP genes regardless of family history or inheritance pattern.

In three simplex RP patients, conventional genotyping and resequencing arrays^{10,12} had failed to detect mutations and high-throughput DNA sequencing was used to look for novel mutations. A novel homozygous mutation was detected in *CNGBI* (c.2957A>T) in a patient presumed initially to have simplex RP (figure 1), but who subsequently reported consanguinity in the family (figure 4). The likely functional effects of this novel homozygous variant (figure 4), absence in controls, segregation with disease and consanguinity in the family support its designation as the causative mutation in this patient. *CNGBI*, which encodes the β -subunit of the rod cGMP-gated channel, is a rare cause of recessive RP (RP45; MIM*600724) with only two mutations previously reported.²¹⁻²³ This mutation would not have been identified by conventional testing strategies because *CNGBI* would have been excluded due to the extremely low frequency of mutations. Therefore, high-throughput DNA sequencing can detect novel variants in addition to any known mutations if applied in previously uncharacterized RP cases. In addition to the immediate benefits for genetic counseling of RP patients high-throughput DNA sequencing provides the opportunity to identify candidates for gene therapy trials.^{7,8}

A diagnostic molecular test must deliver a reliable, cost-effective mutational screen which can be performed in a clinical laboratory. Perhaps the most important factor to ensure detection of variants is sufficient coverage. A similar approach for screening of genes involved in RP is being developed by Daiger SP *et al*²⁴ to target complete genes and additional candidate genes. By focusing on the exonic regions of known RP genes our target region was kept to 359kb, which

although well below the capacity of the capture array, meant that coverage was maximized. The minimal coverage of 15-fold recommended for diagnostic implementation of this technology by Hoischen *et al*²⁵ was achieved in 99% of the target region. We demonstrated that sufficient coverage (>15 fold in 90% of the target region) was maintained when four samples, each labeled with a 4 base pair tag or 'bar code' were pooled and sequenced together. Additional cost savings could be achieved by multiplexing of more samples in a first pass screen and if any did not meet specified coverage criteria or no pathogenic variant was detected, further sequencing could be performed to increase coverage. Whilst high-throughput DNA sequencing might have to be outsourced by smaller diagnostic laboratories the data analysis process performed by our group could be readily adopted by any laboratory.

The sheer length of sequence interrogated by targeted capture and massively parallel sequencing inevitably leads to the detection of many sequence variants and the assessment of which of these are potentially pathogenic is a significant challenge. In this study we have focused on the 'low hanging fruit' represented by missense (or nonsense) variants and assessed pathogenicity from predicted functional effects, frequency in controls and segregation with disease. The successful detection of known and novel mutations justifies the introduction of a mutational screen using this approach. More sophisticated analyses are possible, although if a novel non-coding or synonymous variant does not have an obvious effect upon splicing then it is difficult to assess its likely functional effect. It would not be feasible to test all new variants in a control population as part of a routine diagnostic service, however the availability of an increasing number of genome sequences (1000 Genomes Project) will make it easier to classify common and rare variants.

The power of high-throughput DNA sequencing is exemplified by studies in which targeted sequencing of all coding regions, termed the ‘exome’, successfully identified mutations in two rare Mendelian disorders^{26,27} and whole-genome sequencing detected the genetic basis of Charcot–Marie–Tooth disease.²⁸ Unfortunately, the cost of achieving sufficient coverage and the challenges of data analysis mean that the application of exome and whole-genome sequencing as a clinical diagnostic tool is remote. However, we have demonstrated that targeted sequence capture of a disease-specific cohort and subsequent high-throughput DNA sequencing can be used as a cost-effective genetic diagnostic tool.

ACKNOWLEDGEMENTS

The authors thank Clive Wolsley for his assistance with retinal electrophysiology, Gareth McKay for assistance with control DNA samples and Justin O’Neill for his work in recruiting the families. This research was supported by the Health and Social Care Northern Ireland R&D of the Public Health Agency (Project grant RRG 4.43).

COMPETING INTERESTS

The authors have no competing interests

FUNDING

This research was supported by the Health and Social Care Northern Ireland R&D of the Public Health Agency (Project grant RRG 4.43).

COPYRIGHT LICENCE STATEMENT

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Group and co-owners or contracting owning societies (where published by the BMJ Group on their behalf), and its Licensees to permit this article (if accepted) to be published in The Journal of Medical Genetics and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence.

REFERENCES

1. Guttmacher AE, Collins FS. Welcome to the genomic era. *N Engl J Med* 2003;**349**(10):996-998.
2. Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. *Transl Res* 2009;**154**(6):277-287.
3. Burke W. Genetic testing. *N Engl J Med* 2002;**347**(23):1867-1875.
4. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet* 2009;**85**(2):142-154.
5. Hartong DT, Berson EL, Dryja TP. Retinitis Pigmentosa. *Lancet*. 2006;**368**(9549):1795-809.
6. Drack AV, Lambert SR, Stone EM. From the laboratory to the clinic: molecular genetic testing in pediatric ophthalmology. *Am J Ophthalmol* 2010;**149**(1):10-17.

7. Cremers FP, Collin RW. Promises and challenges of genetic therapy for blindness. *Lancet* 2009;**374**(9701):1569-1570.
8. Smith AJ, Bainbridge JW, Ali RR. Prospects for retinal gene replacement therapy. *Trends Genet* 2009;**25**(4):156-165.
9. Fishman GA. Retinitis pigmentosa. Genetic percentages. *Arch Ophthalmol* 1978;**96**(5):822-826.
10. Koenekoop RK, Lopez I, den Hollander AI, Allikmets R, Cremers FP. Genetic testing for retinal dystrophies and dysfunctions: benefits, dilemmas and solutions. *Clin Experiment Ophthalmol* 2007;**35**(5):473-485.
11. Mandal MN, Heckenlively JR, Burch T, Chen L, Vasireddy V, Koenekoop RK, Sieving PA, Ayyagari R. Sequencing arrays for screening multiple genes associated with early-onset human retinal degenerations on a high-throughput platform. *Invest Ophthalmol Vis Sci* 2005;**46**(9):3355-3362.
12. Clark GR, Crowe P, Muszynska D, O'Prey D, O'Neill J, Alexander S, Willoughby CE, McKay GJ, Silvestri G, Simpson DA. Development of a Diagnostic Genetic Test for Simplex and Autosomal Recessive Retinitis Pigmentosa. *Ophthalmology* 2010. Epub ahead of print (doi:10.1016/j.optha.2010.02.029).
13. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009;**55**(4):641-658.
14. Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med* 2009;**1**(10):100.

15. Chou LS, Liu CS, Boese B, Zhang X, Mao R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* 2010;**56**(1):62-72.
16. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res* 2001;**11**(10):1725-1729.
17. den Hollander AI, Davis J, van der Velde-Visser SD, Zonneveld MN, Pierrottet CO, Koenekoop RK, Kellner U, van den Born LI, Heckenlively JR, Hoyng CB, Handford PA, Roepman R, Cremers FP. CRB1 mutation spectrum in inherited retinal dystrophies. *Hum Mutat.* 2004;**24**:355-69.
18. Bascom RA, Liu L, Humphries P, Fishman GA, Murray JC, McInnes RR. Polymorphisms and rare sequence variants at the ROM1 locus. *Hum Mol Genet* 1993;**2**(11):1975-1977.
19. Berson EL, Grimsby JL, Adams SM, McGee TL, Sweklo E, Pierce EA, Sandberg MA, Dryja TP. Clinical features and mutations in patients with dominant retinitis pigmentosa-1 (RP1). *Invest Ophthalmol Vis Sci* 200;**42**(10):2217-2224.
20. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* 2008;**10**(6):484-492.
21. Bareil C, Hamel CP, Delague V, Arnaud B, Demaille J, Claustres M. Segregation of a mutation in CNGB1 encoding the beta-subunit of the rod cGMP-gated channel in a family with autosomal recessive retinitis pigmentosa. *Hum Genet* 2001;**108**(4):328-334.
22. Kondo H, Qin M, Mizota A, Kondo M, Hayashi H, Hayashi K, Oshima K, Tahira T, Hayashi K. A homozygosity based search for mutations in patients with autosomal

- recessive retinitis pigmentosa, using microsatellite markers. *Invest Ophthalmol Vis Sci* 2004;**45**(12):4433-4439.
23. Becirovic E, Nakova K, Hammelmann V, Hennel R, Biel M, Michalakis S. The retinitis pigmentosa mutation c.3444+1G>A in CNGB1 results in skipping of exon 32. *PLoS One* 2010;**5**(1):e8969.
24. Daiger SP, Sullivan LS, Bowne SJ, Birch DG, Heckenlively JR, Pierce EA, Weinstock GM. Targeted high-throughput DNA sequencing for gene discovery in retinitis pigmentosa. *Adv Exp Med Biol* 2010;**664**:325-331.
25. Hoischen A, Gilissen C, Arts P, Wieskamp N, van der Vliet W, Vermeer S, Steehouwer M, de Vries P, Meijer R, Seiquerios J, Knoers NV, Buckley MF, Scheffer H, Veltman JA. Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum Mutat* 2010 Feb;**31**(4):494-499.
26. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;**461**(7261):272-276.
27. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;**42**(1):30-35.
28. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;**362**(13):1181-1191.

29. Molokanova E, Krajewski JL, Satpaev D, Luetje CW, Kramer RH. Subunit contributions to phosphorylation-dependent modulation of bovine rod cyclic nucleotide-gated channels. *J Physiol* 2003;**552**(Pt 2):345-356.

FIGURE LEGENDS

Figure 1. Fundal images and pedigrees of the five retinitis pigmentosa patients. The fundal appearances of the patients are typical of RP, showing variable amounts of optic disc pallor, attenuation of the retinal vessels, retinal pigment epithelial atrophy, and bone spicule pigmentation.

Figure 2. Sequence coverage of targeted regions. The graph shows the ‘completeness’ of total coverage, i.e. the percentage of the targeted sequences which are represented at different minimum fold coverages. The solid line denotes the sample analyzed in a single lane of a Genome Analyzer flow cell and the dashed line the four samples multiplexed together (mean values \pm SD).

Figure 3. Detection of previously identified mutations in two retinitis pigmentosa patients.

A. Detection of *PDE6B* homozygous mutation c.1685G>A (p.G562D). The target regions (black boxes), across which capture probes were designed, coincide with the exons of the *PDE6B* gene. The target region for exon 13 is expanded and can be seen to extend 100bp 5' from the exon (due to the proximity of exons 13 and 14 the whole 3' intron is captured). Part of the exon is expanded to show the DNA sequence and below this the reads aligned at this position are shown

in red or green to indicate forward and reverse reads (for legibility not all aligned reads are shown). All the sequence reads have an A in place of a C in the reference sequence indicating a homozygous change. This variant is confirmed by Sanger sequencing (bottom) and results in a missense substitution of glycine (GGC) with aspartic acid (GAC). **B.** Detection of *CRBI* compound heterozygous mutations c.2129A>T; p.E710V and c.2548G>A; p.G850S. The targeted regions are depicted as in A. However, for both variants approximately half of the reads match the reference sequence while half carry the variant nucleotide indicating a heterozygous variant.

Figure 4. Detection of a novel mutation (c.2957A>T; p.N986I) in cyclic nucleotide gated channel beta 1 gene (*CNGB1*; MIM*600724). **A.** Alignment of sequence reads to the *CNGB1* reference sequence. All of the 134 aligned reads are depicted as lines below the sequence of *CNGB1* exon 29. A small number of reads are expanded to show their DNA sequence and the T which varies from the reference (A) in all reads displayed, indicating a homozygous mutation c.2957A>T in exon 29 of *CNGB1*. **B.** Pedigree showing the relatedness of the parents of the index case uncovered following the identification of a recessive mutation. The heterozygous state (+/-) of the other available family members indicates segregation of the mutation with disease. **C.** The *CNGB1* protein (NP_001288) contains an ion transport domain (pfam00520) and a cyclic nucleotide-binding domain (CNBD: cd0038; NCBI Conserved Domain Database). The novel mutation results in a non-conservative substitution (p.N986I) of a polar, uncharged, hydrophilic (blue) residue (asparagine: N) for a non-polar, hydrophobic (red) residue (isoleucine: I). **D.** ClustalW alignment of *CNGB1* orthologs from 10 species, with amino acids colored

according to hydrophobicity, reveals the high level of conservation of this region. The asparagine (N986) is an invariant residue in a key structural domain: the cyclic nucleotide-binding domain (CNBD). CNBDs are present in cAMP- and cGMP-dependent protein kinases (cAPK and cGPK) and vertebrate cyclic nucleotide-gated ion-channels. Closure of cyclic nucleotide-gated channels occurs when the concentration of cGMP decreases in response to the G-protein-coupled cascade initiated by light. The decrease in cGMP is thus transduced into membrane hyperpolarization. Previously a homozygous missense mutation in exon 30 of the *CNGB1* gene (c. 2978G>T; p.G993V) was identified as a rare cause of recessive RP in a consanguineous French family²²: the residue is indicated (V). Both of these missense mutations in *CNGB1* result in amino acid substitutions (p.G993V; p.N986I) in the CNBD and may affect the binding of cGMP to CNGB1. Modulation of the sensitivity of CNGB1-containing channels is potentially important for light or dark adaptation and can be achieved by phosphorylation of a tyrosine residue (Y983, equivalent to Y1097 in bovine CNGB1²⁹). The p.N986I mutation is only three residues from this tyrosine (P) and is likely to disrupt phosphorylation by altering the phosphorylase binding motif or changing the local 3D protein structure. The loss of 170 amino acids of the distal C-terminus caused by the only other reported mutation in *CNGB1* (c.3444 + 1G-->A)^{22,23} is indicated by a black line.

Licence for Publication

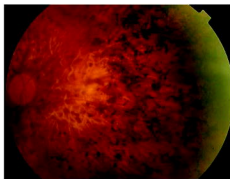
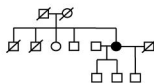
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in

Journal of Medical Genetics and any other BMJPG L products and sublicences such use and exploit all subsidiary rights, as set out in our licence

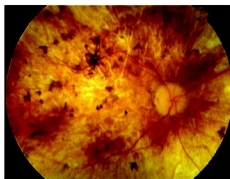
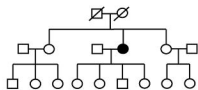
(<http://group.bmj.com/products/journals/instructions-for-authors/licence->

forms)

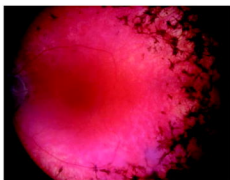
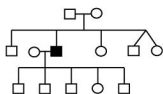
RP75



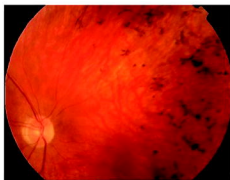
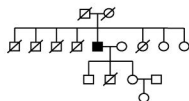
RP141



RP148



RP167



RP179

