

Rapid and cost-effective molecular diagnosis using exome sequencing of one proband with autosomal dominant congenital cataract

J-H Chen^{1,2}, J Qiu¹, H Chen^{1,2}, CP Pang^{1,2}
and M Zhang¹

LABORATORY STUDY

Abstract

Purpose Due to high genetic heterogeneity, to exclude known mutations and map novel mutations in autosomal dominant congenital cataract (ADCC) using conventional candidate gene screening requires laborious laboratory work. We attempted to use a cost-effective exome sequencing strategy to identify disease-causing mutations in an ADCC pedigree.

Methods An ADCC pedigree affected by nuclear cataract and 200 unrelated senile cataract controls were recruited and given comprehensive ophthalmic examination. Whole exome of the proband of the family was captured by the Illumina TruSeq Exome Enrichment Kit, followed by sequencing using Illumina HiSeq 2000 sequencer. Validation was performed by direct sequencing.

Results The whole exome, including all exons of known ADCC disease-causing genes, was screened for possible disease-causing mutations. A recurrent missense mutation c.773C>T (p.S258F) in exon 2 of the gap junction protein alpha 8 gene (*GJA8*) was identified in the proband with nuclear cataract. The result was confirmed by direct sequencing. The mutation showed complete co-segregation with the disease phenotype in the family but was not observed in unrelated unaffected controls.

Conclusion By successfully sequencing whole exome of only one proband and identifying a *GJA8* mutation in one ADCC

pedigree, the current study demonstrated that exome sequencing could serve as a rapid, robust, and cost-effective approach in clinical diagnosis and disease-causing gene discovery for ADCC.

Eye (2014) 28, 1511–1516; doi:10.1038/eye.2014.158; published online 10 October 2014

Introduction

Autosomal dominant congenital cataract (ADCC) is the most common form of inherited cataracts.¹ Albeit advances in surgical management and effective visual rehabilitation of most cataract patients, early diagnosis is important to optimize visual outcome in cataract surgeries.²

Linkage and candidate gene studies have found high genetic heterogeneities in ADCC,^{3–6} for which at least 18 associated genes have been identified. To exclude known mutations and map novel mutations in all ADCC disease-causing genes using conventional gene screening strategy requires laborious laboratory work, and it is impractical for genetic diagnosis and testing in clinical practice. Exome sequencing has recently been successfully applied in mapping disease-causing genes and mutations for Mendelian diseases and tumors.^{7,8} By using exome capture chips followed by next-generation sequencing, the technique exhaustively sequences and identifies potential structural or functional variants in the exome.

¹Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, China

²Department of Ophthalmology and Visual Sciences, the Chinese University of Hong Kong, Hong Kong, China

Correspondence: M Zhang, Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, North Dongxia Road, Shantou, Guangdong 515041, China
Tel: +86 754 88393560;
Fax: +86 754 88393501.
E-mail: zmz@jsiec.org

Received: 18 April 2013
Accepted in revised form: 2 March 2014
Published online: 10 October 2014

In the current study, we attempt to evaluate a new approach of ADCC genetic study using exome sequencing. The whole exome was sequenced for one proband from a pedigree with autosomal dominant congenital nuclear cataract. By successfully identifying a recurrent mutation in the proband, the current study demonstrated a new, rapid, and cost-effective approach to identify disease-causing mutations in ADCC.

Materials and methods

Sample selection and clinical examination

This study was approved by the Ethics Committee of Joint Shantou International Eye Center and was conducted in accordance with the Declaration of Helsinki. Written consent was obtained from each participating subject after explanation of the nature of the study.

A Han Chinese family with autosomal dominant congenital nuclear cataract (Figure 1) and 200 unrelated controls with mild senile cataracts were collected at the Joint Shantou International Eye Center, Shantou, China. Comprehensive ophthalmic examinations were performed in all participants. Photographs of the anterior segment were taken from ADCC patients.

Exome sequencing of the proband

Peripheral blood was collected from the proband (III-8) and other family members and all senile cataract controls. Genomic DNA was extracted by using the QIAmp Blood kit (Qiagen, Hilden, Germany). Genomic DNA (3 μg) was used to perform exome sequencing by using service provided by Axseq Technologies, Inc. (Rockville, MD, USA). Whole exome of the proband was captured by Illumina

TruSeq Exome Enrichment Kit (Illumina, San Diego, CA, USA), and was used to prepare libraries and cluster using Illumina HiSeq 2000, and sequenced with paired-end 100 base length configuration.

Bioinformatics analysis of exome data

The reads are mapped against UCSC hg19 Human Reference Genome (<http://genome.ucsc.edu/>) by using BWA (<http://bio-bwa.sourceforge.net/>). The single-nucleotide variations (SNVs) and Indels are detected by SAMTOOLS (<http://samtools.sourceforge.net/>), and previously known and reported variants were identified and filtered using dbSNP 135 (<http://www.ncbi.nlm.nih.gov/snp/>) and 1000 genome project (<http://www.1000genomes.org>) data. A step-by-step filtering strategy was used to select candidates of disease-causing mutations. Functional annotation was performed using ANNOVAR (<http://www.openbioinformatics.org/annovar/>), and coding consequence predictions for nonsynonymous SNVs and indels were done using Sorting Tolerant From Intolerant (SIFT; <http://sift.bii.a-star.edu.sg/index.html>), PolyPhen (<http://genetics.bwh.harvard.edu/pph/>), and Condel (<http://bg.upf.edu/fannsd/>). A total of 18 previously reported ADCC disease-causing genes were selected to analyze: night crystallin genes, including αA-crystallin (CRYAA),⁹ αB-crystallin (CRYAB),¹⁰ βA1-crystallin (CRYBA1),¹¹ βA4-crystallin (CRYBA4),¹² βB1-crystallin (CRYBB1),¹³ βB2-crystallin (CRYBB2),¹⁴ γC-crystallin (CRYGC),¹⁵ γD-crystallin (CRYGD),¹⁶ and γS-crystallin (CRYGS);¹⁷ one cytoskeletal protein gene: beaded filament structural protein 2 (BFSP2);¹⁸ three membrane protein genes: gap junction protein alpha 3 (GJA3),¹⁹ gap junction protein alpha 8 (GJA8),²⁰ and major intrinsic protein of lens fiber

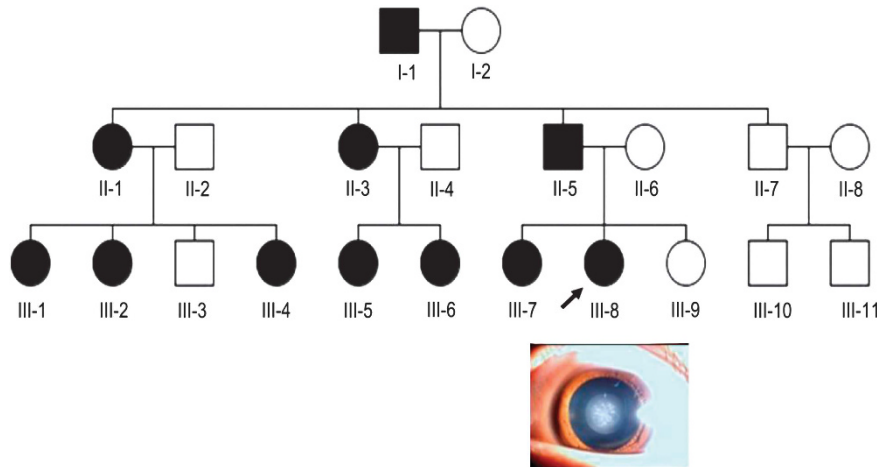


Figure 1 A Chinese pedigree with congenital nuclear cataract. Filled squares and circles denote affected males and females, respectively. Normal individual are shown as empty symbols. The arrow denotes the proband, with his photo of slit-lamp examination shown below.

(MIP);²¹ three growth and transcription factor genes, including heat-shock transcription factor 4 (*HSF4*),²² paired-like homeodomain 3 (*PITX3*),²³ and v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog (*MAF*);²⁴ and two other genes, including chromatin modifying protein-4B (*CHMP4B*),²⁵ and Ephrin receptor A2 (*EPHA2*).²⁶ For splice site variants, indels, and nonsynonymous SNVs predicted to be functionally deleterious, a comprehensive literature search was conducted to identify known disease-causing variants of ADCC. Candidate mutations in the 18 ADCC genes that were identified by bioinformatic analysis were selected for further validation. The same strategy was applied to analyze additional 212 genes that are recorded in the Cat-Map database (<http://cat-map.wustl.edu/>),²⁷ which contained >20 known genes identified for non-syndromic cataract, >130 genes and loci for syndromic forms of cataract, and >70 mouse mutants with cataract.

Variant validation using Sanger sequencing

Genomic sequences of the genes containing candidate variants selected by bioinformatic analysis were obtained from the NCBI reference sequence database (<http://www.ncbi.nlm.nih.gov/refseq>). Primers designed accordingly by Primer 3 were used to amplify coding exons containing the candidate variants and their flanking regions (GJA8-1F 5'-AATGTGGTGGACT GCTTCGT-3' and GJA8-1R 5'-GGCAGTGTCTCTTGGA GCC-3'). The PCR amplification was performed as previously described.^{28,29} Bidirectional sequencing of PCR products was performed using the BigDye Terminator Cycle Sequencing v3.1 kit (ABI, Foster City, CA, USA) and the 3130xl Genetic Analyzer (ABI) following the protocol suggested by the manufacturer. Sequence alignment and analysis of variations were performed using NovoSNP.³⁰

Results

Statistics of exome sequencing data

The total sequencing depth in the proband is shown in Table 1. In all, 87.6% target region was covered with $\geq 10\times$ sequencing depth, and 94.20% was covered with at least $1\times$ sequencing depth. The median and mean read depths of target regions were $51.0\times$ and $54.9\times$, respectively, which should ensure high quality of SNV and indel calling.

SNVs and indels in known ADCC disease-causing genes

Exome sequencing identified a total of 19 347 coding SNV and 475 coding indels in the proband (Table 2). All of the

Table 1 Summary of exome sequencing data in the proband

| | Statistics |
|--|---------------|
| Total reads | 68 314 400 |
| Total yield (bp) | 6 899 754 400 |
| Average read length (bp) | 101 |
| Mappable reads (= reads mapped to human genome) | 61 093 660 |
| Mappable yield (bp) | 6 065 274 734 |
| % Mappable reads (out of total reads) | 89.40% |
| Mean read depth | $111\times$ |
| On-target reads (= reads mapped to target regions) | 41 707 646 |
| On-target yield (bp) | 3 410 856 645 |
| % On-target reads (out of mappable reads) | 68.30% |
| Target regions (bp) | 62 085 286 |
| % Coverage of target regions ($>1\times$) | 94.20% |
| Number of on-target genotypes ($>1\times$) | 58 458 473 |
| % Coverage of target regions ($>10\times$) | 87.60% |
| Number of on-target genotypes ($>10\times$) | 54 370 183 |
| Median read depth of target regions | $51.0\times$ |
| Mean read depth of target regions | $54.9\times$ |

Table 2 Filtering of SNVs and indels in the proband by bioinformatics analysis and experimental validation

| Sample name | SNV | Indel |
|--|--------|--------|
| Total number | 68 845 | 14 352 |
| Coding variants | 19 347 | 475 |
| In 18 ADCC disease-causing genes | 40 | 0 |
| Not in dbSNP or 1000 Genome database | 7 | 0 |
| Nonsynonymous SNVs or indels | 2 | 0 |
| Predicted to be deleterious | 2 | 0 |
| Known disease-causing mutations | 1 | 0 |
| Co-segregation with ADCC phenotype in the pedigree | 1 | 0 |

exons in the 18 ADCC disease-causing genes in the proband were covered in the current study. A total of 48 coding SNVs and no coding indel were observed in the known ADCC genes. Among these variants, a previously reported heterozygous missense mutation c.773C>T (p.S258F) was observed in exon 2 of *GJA8*, which was predicted by bioinformatics analysis to be functionally deleterious.²⁰ The mutation was further confirmed in the proband by direct sequencing. It was also found in all other affected family members but was absent in either unaffected family members or 200 unrelated senile cataract controls. It thus completely co-segregated with the ADCC disease phenotype (Figure 2). Another nonsynonymous SNV c.49G>A (p.V17M) in exon3 of *HSF* was also observed in the proband. However, it was not detected in other affected family members and hence did not show co-segregation with the disease phenotype. We have also further analyzed additional 212 genes from the Cat-Map database that had been reported to

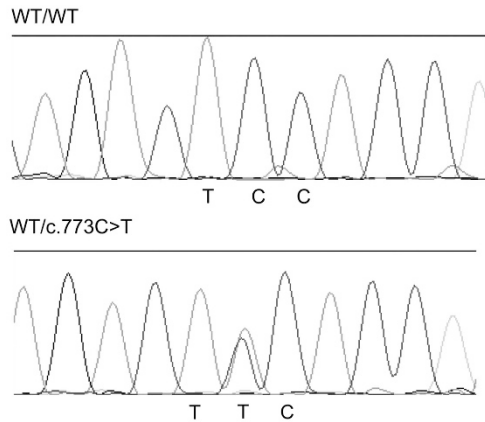


Figure 2 Sanger sequencing confirms the disease-causing mutation in *GJA8* that was identified by exome sequencing. The upper panel shows the chromatography of unaffected family members or senile cataract controls; the lower panel shows the chromatography of the heterozygous mutation *c.773C>T* (p.S258F) in exon2 of *GJA8*.

be associated with cataract phenotype in humans or animals in the literature. No other variant was found to be co-segregated with the ADCC disease phenotype (Supplementary Table S1).

Discussion

In this study, we used an exome sequencing approach to screen mutations in a pedigree with ADCC. We performed exome sequencing in the proband of the pedigree and successfully identified a previously reported recurrent disease-causing mutation and thus demonstrated the robustness and potential application of exome sequencing in clinical molecular diagnosis of ADCC.

By sequencing the whole exome of one proband in each ADCC pedigree, the approach was able to screen all coding variants in 18 ADCC disease-causing genes in a rapid and cost-effective way. In the current study a heterozygous missense mutation in *GJA8* was identified, demonstrating the potential application of exome sequencing in the molecular diagnosis of ADCC. One important limiting factor in clinical application of molecular diagnosis is cost. Exons constitute approximately 1% of human genome, and coding variants in these genomic regions are more likely to be functionally relevant and probably affect phenotypes or diseases, especially for Mendelian diseases such as ADCC. By targeting these specific regions in the human genome, exome sequencing provides a more cost-effective method to map genetic variants with large effect, compared with whole-genome sequencing.

The cost of next-generation sequencing has been reduced dramatically recently, especially for exome sequencing, which allows possible applications of exome sequencing in molecular diagnosis. The cost of exome sequencing with $100\times$ average sequencing depth can be less than one tenth of that of whole genome sequencing with $30\times$ average sequencing depth.³¹ Therefore, with a fixed budget, exome sequencing can achieve higher sequencing depth in functionally relevant genomic region and thus significantly increases power of detection of disease-causing mutations. Another limiting factor of molecular diagnosis methods is turnaround time. In our study, the single sequencer (Illumina HiSeq 2000) generated 600 million pair-end reads of length 100 in about 11 days.³² This sequencing process can probably be shortened to about 17 h when using the new model Illumina HiSeq 2500 to do the whole-exome sequencing.³³ By focusing on 18 ADCC genes, the completion of bioinformatics analysis could be controlled within only 5 h. The total turnaround time is feasible for clinical molecular diagnosis.

Congenital cataract is of high genetic heterogeneities. To exclude known mutations and map novel mutations in all ADCC disease-causing genes using such conventional strategy requires extensive laboratory work and budget, and it is especially impractical for genetic diagnosis and testing in clinical practice. In addition, in the current study we focused on 18 genes that were previously reported to cause ADCC. The panel can be extended to include additional new disease-causing genes of ADCC. Our findings also showed that by using a high sequencing depth strategy exome sequencing of only one proband or affected family member was able to detect the underlying disease-causing mutation in an ADCC pedigree. The coverage of target regions (with $\geq 10\times$ sequencing depth) was 87.6%; the average sequencing depth was $111\times$; and the average sequencing depth on target region was $51\times$. In general, an average sequencing depth $\geq 100\times$ could ensure high coverage of target regions and high quality of SNV and indel calling. Our data showed that all 18 ADCC genes were covered in the exome sequencing. Bioinformatics analysis could be very powerful in identifying candidate variants in these genes, which could be quickly validated in ADCC pedigrees to confirm the possible co-segregation with the disease phenotype. The approach can also be useful in search for unknown ADCC genes. If no disease-causing mutations are found among these genes in the affected members of the pedigree, the results implicated novel disease-causing genes in the ADCC pedigrees.

In the current study, we proposed and evaluated an approach of ADCC genetic study to sequence whole

exome of one proband from a pedigree with autosomal dominant congenital nuclear cataract. By successfully identifying a known mutation in the proband, the current study demonstrated a rapid and cost-effective approach to map disease-causing mutations in ADCC.

Summary

What was known before

- Due to high genetic heterogeneity, to exclude known mutations and map novel mutations in autosomal dominant congenital cataract (ADCC) using conventional candidate gene screening requires laborious laboratory work.

What this study adds

- By successfully sequencing whole exome of only one proband and identifying a *GJA8* mutation in one ADCC pedigree, the current study demonstrated that exome sequencing could serve as a rapid, robust, and cost-effective approach in clinical diagnosis and disease-causing gene discovery for ADCC.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This study was supported in part by the research grant from National Natural Science Foundation of China (Nos. 81000397, 81371033, 30901646, and 81170853), Natural Science Foundation of Guangdong Province, China (No. S2013010015618), Science and Technology Planning Project of Guangdong Province, China (No. 2010B031600130 and 2011B031300013), and Joint Shantou International Eye Center, Shantou University/The Chinese University of Hong Kong (Nos. 10-019, 10-020, 10-021, and 10-022). We would like to thank Ms Elaine To and Dr Lijia Chen for help with preparing the manuscript and Ms Yuqian Zheng and Ms Ciyan Xu for technical support.

References

- 1 Bateman JB, Geyer DD, Flodman P, Johannes M, Sikela J, Walter N *et al*. A new betaA1-crystallin splice junction mutation in autosomal dominant cataract. *Invest Ophthalmol Vis Sci* 2000; **41**(11): 3278–3285.
- 2 Dawson CR, Schwab IR. Epidemiology of cataract—a major cause of preventable blindness. *Bull World Health Organ* 1981; **59**(4): 493–501.
- 3 Zhang L-Y, Yam GH-F, Fan DS-P, Tam PO-S, Lam DS-C, Pang C-P. A novel deletion variant of gammaD-crystallin responsible for congenital nuclear cataract. *Mol Vis* 2007; **13**: 2096–2104.
- 4 Zhang L-Y, Gong B, Tong J-P, Fan DS-P, Chiang SW-Y, Lou D *et al*. A novel gammaD-crystallin mutation causes mild changes in protein properties but leads to congenital coralliform cataract. *Mol Vis* 2009; **15**: 1521–1529.
- 5 Li Y, Wang J, Dong B, Man H. A novel connexin46 (GJA3) mutation in autosomal dominant congenital nuclear pulverulent cataract. *Mol Vis* 2004; **10**: 668–671.
- 6 Wang K, Wang B, Wang J, Zhou S, Yun B, Suo P *et al*. A novel GJA8 mutation (p.I31T) causing autosomal dominant congenital cataract in a Chinese family. *Mol Vis* 2009; **15**: 2813–2820.
- 7 Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM *et al*. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010; **42**(1): 30–35.
- 8 Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA *et al*. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* 2010; **330**(6009): 1410–1413.
- 9 Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW, Weleber RG. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet* 1998; **7**(3): 471–474.
- 10 Chen Q, Ma J, Yan M, Mothobi ME, Liu Y, Zheng F. A novel mutation in CRYAB associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Mol Vis* 2009; **15**: 1359–1365.
- 11 Kannabiran C, Rogan PK, Olmos L, Basti S, Rao GN, Kaiser-Kupfer M *et al*. Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene. *Mol Vis* 1998; **4**: 21.
- 12 Zhou G, Zhou N, Hu S, Zhao L, Zhang C, Qi Y. A missense mutation in CRYBA4 associated with congenital cataract and microcornea. *Mol Vis* 2010; **16**: 1019–1024.
- 13 Mackay DS, Boskovska OB, Knopf HL, Lampi KJ, Shiels A. A nonsense mutation in CRYBB1 associated with autosomal dominant cataract linked to human chromosome 22q. *Am J Hum Genet* 2002; **71**(5): 1216–1221.
- 14 Vanita, Sarhadi V, Reis A, Jung M, Singh D, Sperling K *et al*. A unique form of autosomal dominant cataract explained by gene conversion between beta-crystallin B2 and its pseudogene. *J Med Genet* 2001; **38**(6): 392–396.
- 15 Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N *et al*. The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet* 1999; **65**(5): 1261–1267.
- 16 Nandrot E, Slingsby C, Basak A, Cherif-Chefchaouni M, Benazzouz B, Hajaji Y *et al*. Gamma-D crystallin gene (CRYGD) mutation causes autosomal dominant congenital cerulean cataracts. *J Med Genet* 2003; **40**(4): 262–267.
- 17 Vanita V, Singh JR, Singh D, Varon R, Sperling K. Novel mutation in the gamma-S crystallin gene causing autosomal dominant cataract. *Mol Vis* 2009; **15**: 476–481.
- 18 Jakobs PM, Hess JF, FitzGerald PG, Kramer P, Weleber RG, Litt M. Autosomal-dominant congenital cataract associated with a deletion mutation in the human beaded filament protein gene BFSP2. *Am J Hum Genet* 2000; **66**(4): 1432–1436.
- 19 Mackay D, Ionides A, Kibar Z, Rouleau G, Berry V, Moore A *et al*. Connexin46 mutations in autosomal dominant congenital cataract. *Am J Hum Genet* 1999; **64**(5): 1357–1364.
- 20 Gao X, Cheng J, Lu C, Li X, Li F, Liu C *et al*. A novel mutation in the connexin 50 gene (GJA8) associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Curr Eye Res* 2010; **35**(7): 597–604.

- 21 Yang G, Zhang G, Wu Q, Zhao J. A novel mutation in the MIP gene is associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Mol Vis* 2011; **17**: 1320–1323.
- 22 Bu L, Jin Y, Shi Y, Chu R, Ban A, Eiberg H *et al*. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nat Genet* 2002; **31**(3): 276–278.
- 23 Semina EV, Ferrell RE, Mintz-Hittner HA, Bitoun P, Alward WL, Reiter RS *et al*. A novel homeobox gene PITX3 is mutated in families with autosomal-dominant cataracts and ASMD. *Nat Genet* 1998; **19**(2): 167–170.
- 24 Vanita V, Singh D, Robinson PN, Sperling K, Singh JR. A novel mutation in the DNA-binding domain of MAF at 16q23.1 associated with autosomal dominant "cerulean cataract" in an Indian family. *Am J Med Genet A* 2006; **140**(6): 558–566.
- 25 Shiels A, Bennett TM, Knopf HL, Yamada K, Yoshiura K, Niikawa N *et al*. CHMP4B, a novel gene for autosomal dominant cataracts linked to chromosome 20q. *Am J Hum Genet* 2007; **81**(3): 596–606.
- 26 Zhang T, Hua R, Xiao W, Burdon KP, Bhattacharya SS, Craig JE *et al*. Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. *Hum Mutat* 2009; **30**(5): E603–E611.
- 27 Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Mol Vis* 2010; **16**: 2007–2015.
- 28 Chen J, Tsang SY, Zhao CY, Pun FW, Yu Z, Mei L *et al*. GABRB2 in schizophrenia and bipolar disorder: disease association, gene expression and clinical correlations. *Biochem Soc Trans* 2009; **37**(Pt 6): 1415–1418.
- 29 Chen JH, Lin W, Sun G, Huang C, Huang Y, Chen H *et al*. A novel PAX6 deletion in a Chinese family with congenital aniridia. *Mol Vis* 2012; **18**: 989–995.
- 30 Weckx S, Del-Favero J, Rademakers R, Claes L, Cruys M, De Jonghe P *et al*. novoSNP, a novel computational tool for sequence variation discovery. *Genome Res* 2005; **15**(3): 436–442.
- 31 Stein LD. The case for cloud computing in genome informatics. *Genome Biol* 2010; **11**(5): 207.
- 32 Liu CM, Wong T, Wu E, Luo R, Yiu SM, Li Y *et al*. SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 2012; **28**(6): 878–879.
- 33 Mason CE, Elemento O. Faster sequencers, larger datasets, new challenges. *Genome Biol* 2012; **13**(3): 314.

Supplementary Information accompanies this paper on Eye website (<http://www.nature.com/eye>)