

# Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing

Mark B. Consugar, MS<sup>1</sup>, Daniel Navarro-Gomez, BS<sup>1</sup>, Emily M. Place, MS<sup>1</sup>, Kinga M. Bujakowska, PhD<sup>1</sup>, Maria E. Sousa, BA<sup>1</sup>, Zoë D. Fonseca-Kelly, PhD<sup>1</sup>, Daniel G. Taub, BA<sup>1</sup>, Maria Janessian, BA<sup>1</sup>, Dan Yi Wang, PhD<sup>1</sup>, Elizabeth D. Au, PhD<sup>1</sup>, Katherine B. Sims, MD<sup>2</sup>, David A. Sweetser, MD<sup>3,4</sup>, Anne B. Fulton, MD<sup>5</sup>, Qin Liu, PhD<sup>1</sup>, Janey L. Wiggs, MD, PhD<sup>1</sup>, Xiaowu Gai, PhD<sup>1</sup> and Eric A. Pierce, MD, PhD<sup>1,6</sup>

**Purpose:** Next-generation sequencing–based methods are being adopted broadly for genetic diagnostic testing, but the performance characteristics of these techniques with regard to test accuracy and reproducibility have not been fully defined.

**Methods:** We developed a targeted enrichment and next-generation sequencing approach for genetic diagnostic testing of patients with inherited eye disorders, including inherited retinal degenerations, optic atrophy, and glaucoma. In preparation for providing this genetic eye disease (GEDi) test on a CLIA–certified basis, we performed experiments to measure the sensitivity, specificity, and reproducibility, as well as the clinical sensitivity, of the test.

**Results:** The GEDi test is highly reproducible and accurate, with sensitivity and specificity of 97.9 and 100%, respectively, for single-nucleotide variant detection. The sensitivity for variant detection was

notably better than the 88.3% achieved by whole-exome sequencing using the same metrics, because of better coverage of targeted genes in the GEDi test as compared with a commercially available exome capture set. Prospective testing of 192 patients with inherited retinal degenerations indicated that the clinical sensitivity of the GEDi test is high, with a diagnostic rate of 51%.

**Conclusion:** Based on quantified performance metrics, the data suggest that selective targeted enrichment is preferable to whole-exome sequencing for genetic diagnostic testing.

*Genet Med* advance online publication 20 November 2014

**Key Words:** genetic diagnostic testing; next-generation sequencing; reproducibility; sensitivity; specificity

## INTRODUCTION

Next-generation sequencing (NGS)–based testing methods are increasingly being used for genetic diagnostic testing. This is especially true for genetically heterogeneous disorders, such as inherited retinal degenerations (IRDs), hearing loss, cardiomyopathies, mitochondrial disorders, and cancer.<sup>1–4</sup> There are multiple advantages to these approaches, including the ability to simultaneously sequence many genes and to quantify allele frequency.<sup>1,3,5,6</sup> The Next-generation Sequencing: Standardization of Clinical Testing workgroup of the Centers for Disease Control and Prevention and the American College of Medical Genetics and Genomics have issued guidelines for clinical laboratory standards for NGS-based testing methods; these include determination of test accuracy, analytical sensitivity and specificity, reproducibility, and repeatability.<sup>7,8</sup> Defining these test characteristics is

important, and despite the increased use of NGS-based tests in CLIA– and/or College of American Pathologists–certified laboratories,<sup>3,6,9</sup> the sensitivity, specificity, and reproducibility of these techniques have been defined for only a small subset of tests.<sup>1,5,10</sup>

Inherited eye disorders are important causes of vision loss. IRDs are among the most common causes of blindness among working-age people,<sup>11</sup> and glaucoma is a leading cause of irreversible blindness worldwide.<sup>12,13</sup> Genetic diagnostic testing for these disorders is challenging because of their genetic heterogeneity. For example, mutations in more than 200 genes can cause IRDs, and multiple genes underlie inherited forms of glaucoma and optic atrophy<sup>14–16</sup> (a complete list of genes is available at <https://sph.uth.edu/retnet/>). Obtaining genetic diagnoses for patients with these disorders is increasingly desired because this information can influence patient care by both informing

The first three authors contributed equally to this work.

<sup>1</sup>Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary and Harvard Medical School, Boston, Massachusetts, USA; <sup>2</sup>Neurogenetics DNA Diagnostic Lab, Department of Neurology, Center for Human Genetics Research, Massachusetts General Hospital, Boston, Massachusetts, USA; <sup>3</sup>Division of Medical Genetics, Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts, USA; <sup>4</sup>Division of Hematology/Oncology, Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts, USA; <sup>5</sup>Department of Ophthalmology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA; <sup>6</sup>Berman-Gund Laboratory, Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA. Correspondence: Eric A. Pierce ([eric\\_pierce@meei.harvard.edu](mailto:eric_pierce@meei.harvard.edu))

Submitted 8 July 2014; accepted 23 October 2014; advance online publication 20 November 2014. doi:10.1038/gim.2014.172

genetic risk assessment and identifying patients who would benefit from novel gene-based therapies.<sup>17–21</sup>

Several groups have recently reported the use of NGS techniques for genetic diagnostic testing of patients with IRDs.<sup>22–26</sup> These reports demonstrate that NGS combined with targeted enrichment approaches is a superior method for genetic diagnostic testing for patients with IRDs in that it improved diagnostic rates and reduced cost compared with traditional sequencing methods. In the reports published to date, however, the accuracy, sensitivity, and specificity, as well as the reproducibility and repeatability of the NGS approaches used have not been completely defined.

We developed a targeted enrichment and NGS approach for genetic diagnostic testing of patients with inherited eye disorders, including IRDs, optic atrophy, and glaucoma. In preparation for providing this genetic eye disease (GEDi) test on a CLIA-certified basis, we sought to determine the accuracy, sensitivity and specificity, reproducibility and repeatability, and clinical sensitivity of this testing approach compared with whole-exome sequencing (WES).

## MATERIALS AND METHODS

### Patient samples

The clinical study was approved by the institutional review boards of the University of Pennsylvania and the Massachusetts Eye and Ear Infirmary and conformed to the tenets of the Declaration of Helsinki. Informed consent was obtained from all participants, who were recruited after having been identified as having a form of IRD following clinical evaluation by investigators E.A.P. or A.B.F. at Massachusetts Eye and Ear Infirmary and Children's Hospital Boston, respectively. Genomic DNA was extracted from patient blood using the PreAnalytiX PAXgene Blood DNA Kit (PAXgene Blood DNA Kit Handbook, October 2009; Qiagen/BD Biosciences, Valencia, CA) or DNAzol (Life Technologies, Carlsbad, CA).

### Targeted enrichment

**Targeted enrichment bait library design.** The custom SureSelect targeted enrichment GEDi capture kit (Agilent Technologies, Santa Clara, CA) was designed to capture and enrich coding exons, 5′-/3′-untranslated regions (UTRs), and select deep intronic regions known to harbor pathogenic mutations associated with the 214 known IRD disease genes described in the Retinal Information Network database (<https://sph.uth.edu/Retnet/>) up to April 2013, as well as eight early-onset glaucoma and optic-atrophy genes, using Agilent Technologies' eArray Web design tool (<https://earray.chem.agilent.com/earray/>). The GEDi capture kit also includes 24 candidate IRD disease genes, 9 age-related macular degeneration risk factor genes, and 1 nonsyndromic hearing loss gene. Additional information regarding the parameters used for GEDi capture kit design is available in the **Supplementary Material** online. The custom mitochondrial genome targeted enrichment baits we designed previously were also included as part of the GEDi capture kit.<sup>27</sup>

Complete lists of the GEDi targeted genes and intronic regions are shown in **Supplementary Table S1** online.

**Capture library sample preparation.** Illumina-compatible paired-end/multiplexable GEDi targeted enrichment capture libraries were generated using a BRAVO automated workstation using standard methods (Agilent methods, part no.: G7530-90000; Protocol v2.1, May 2011) and the following parameters: (i) no less than 1.5 μg of sheared genomic DNA was used for precapture library generation; (ii) five cycles of precapture polymerase chain reaction were used for all samples; (iii) no less than 400 ng of precapture library was used during bait hybridizations; (iv) 14 cycles of postcapture polymerase chain reaction were used to generate all capture libraries; (v) all samples were indexed after capture; and (vi) sample multiplex ratios were determined based on sample concentrations indexed after capture (Agilent methods, part no. G7530-90000; Protocol v2.1, May 2011).

### NGS analysis

GEDi targeted enrichment samples were sequenced using a MiSeq NGS platform (Illumina, San Diego, CA). A 12× patient sample multiplex was clustered to an average cluster density of between 750 and 900 K clusters/mm<sup>2</sup> and 121×6×121 base pairs (bp) indexed/paired-end analyzed using Illumina's 300 cycle MiSeq Reagent Kit V2.

### Whole-exome sequencing

Whole-exome capture and sequencing were performed as described in the **Supplementary Methods** online.

### Informatics analyses

The sequence data obtained were analyzed using a combination of publically available and custom software tools, as described elsewhere.<sup>28</sup> Briefly, BWA (version 0.6.2-r126) was used to align the sequence reads to the human reference genome used by the 1000 Genomes Project. SAMtools (version 0.1.18 or r982:295) was used to remove potential duplicates and make initial single-nucleotide polymorphism (SNP) and insertion/deletion (indel) calls, which were refined using a custom program.<sup>28</sup> A coverage depth cutoff of 10× was applied. Resulting variant calls were annotated using our custom human bp codon resource.<sup>28</sup> Custom scripts were also developed and used to identify candidate variants that fit different filtering criteria, such as genetic models. Variants that fit the appropriate inheritance patterns, and that were rare based on data from the 1000 Genomes Project; the National Heart, Lung, and Blood Institute Exome Sequencing Project Exome Variant Server; and our own internal controls, were considered to be potentially pathogenic. See the **Supplementary Methods** online for additional information.

## RESULTS

### GEDi test design

Probes were designed for 257 genes targeted by the GEDi selective capture system as well as the mitochondrial genome

because retinal degeneration and optic atrophy can accompany mitochondrial disease.<sup>27</sup> Probes for previously identified deep intronic mutations in *CEP290*, *OFD1*, and *USH2A* were also included in the GEDi probe set.<sup>29–31</sup> The targeted regions constitute 1,210,190 base pairs in total (703,980-bp coding sequence) and are listed in **Supplementary Table S1** online.

Probes for some of the targeted regions could not be designed because of the presence of repetitive or nonunique sequence elements. In total, there were 688 such design gaps, ranging from 1 to 2,031 bp in length, with an average length of 112 bp, accounting for a total of 76,980 bp (9,220-bp coding sequence). Analysis of empiric GEDi data shows that design gaps  $\leq 75$  bp (67% of gaps) were relatively well covered by “near-target” capture (**Figure 1a**).

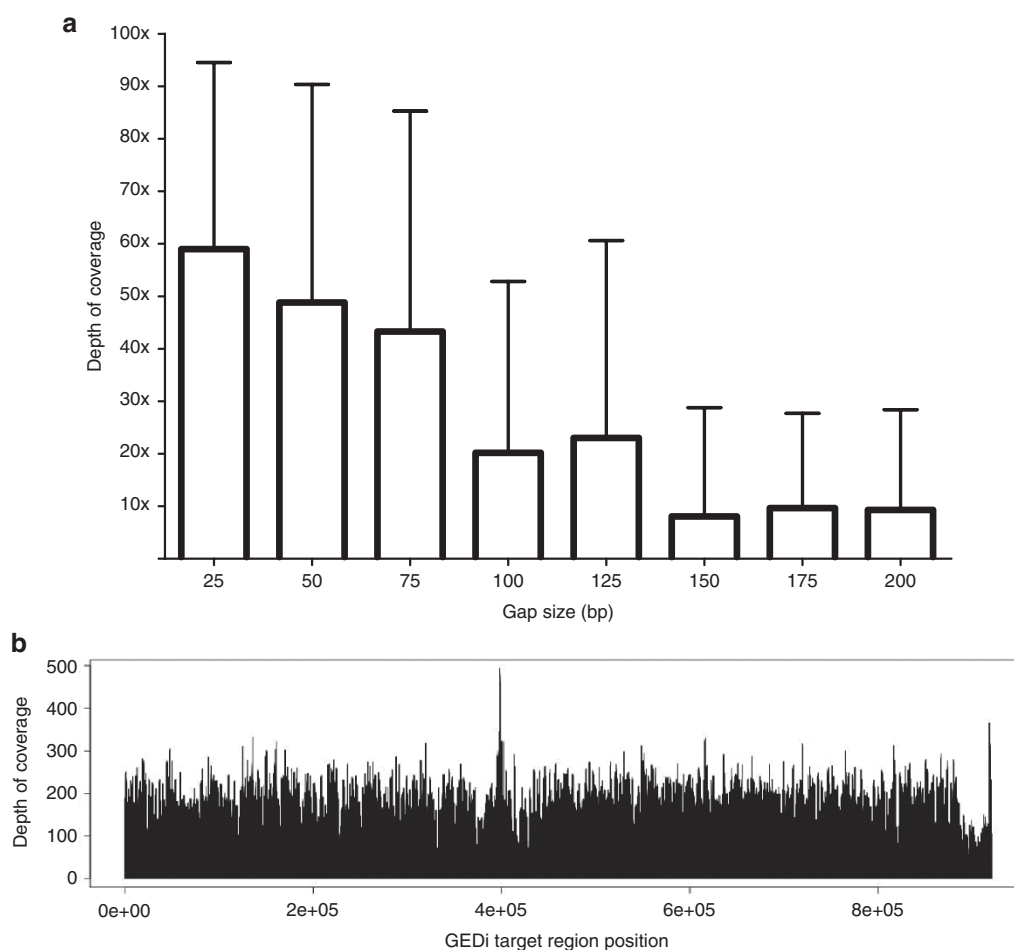
### NGS metrics

**Figure 1b** shows a representative depth-of-coverage (DoC) plot for a 12 $\times$ -multiplexed sample captured using the GEDi targeted enrichment kit and sequenced using an Illumina MiSeq. The data show relatively uniform coverage of the target regions. The average percentages of the target regions covered

at 1 $\times$  (99.8%), 10 $\times$  (98.6%), and 20 $\times$  (96.4%) DoC were also relatively constant for all of the sequencing analyses. The 1.4% of target regions that were not covered with  $\geq 10\times$  read depth included part or all of 14 exons. The overall average DoC for all samples analyzed was  $98.8 \pm 14.5\times$ .

### Test performance metrics

The Next-generation Sequencing: Standardization of Clinical Testing workgroup and the American College of Medical Genetics and Genomics recommend that validation of an NGS-based diagnostic test include performance-test characteristics for assay accuracy, analytical sensitivity and specificity, and reproducibility and repeatability.<sup>7,8</sup> To measure these parameters for the GEDi capture and sequencing test, four samples (three randomly selected patient samples and the NA12878 HapMap sample) were prepared and sequenced in triplicate on each of three separate days. We also performed WES and SNP array genotyping analyses of these four samples using an Agilent V4+UTR whole-exome enrichment kit and Illumina Omni 2.5 SNP arrays, respectively (see **Supplementary Methods** online). The HapMap



**Figure 1 Sequencing metrics.** (a) Analysis of empiric genetic eye disease (GEDi) data shows that design gaps  $\leq 75$  base pairs (bp) were relatively well covered by “near-target” capture. (b) Representative depth-of-coverage (DoC) plot for a 12 $\times$ -multiplexed capture sample using the GEDi targeted enrichment kit and 2  $\times$  121 bp paired-end sequenced using an Illumina MiSeq.

sample was included as an internal control for establishing quality-control metrics and is included in all diagnostic runs to evaluate each diagnostic capture and sequencing run.

**Sensitivity and specificity**

To assess the sensitivity and specificity of the GEDi test, we used the 2,443 SNPs located in GEDi genes that are represented on the Omni 2.5 SNP array, using the Omni 2.5 data as the “gold standard.” For these analyses, sensitivity was calculated as the ability of the GEDi test to correctly identify a SNP when it was identified in the Omni 2.5 data. Similarly, specificity was calculated as the ability of the GEDi test to correctly identify the lack of a variant at a given position when a reference was detected by the Omni 2.5 array<sup>5</sup> (Table 1). For example, 495 ± 1 SNPs identified in the nine GEDi replicates for the OGI-132–357 sample (range, 492–497) were also identified in the Omni 2.5 data, and these were scored as true positives (Table 1). The GEDi test did not identify variants at 10 ± 1.4 positions where variants were identified in the Omni 2.5 data for OGI-132–357, and these were scored as false negatives, giving a sensitivity of 0.98 for variant detection. The GEDi test did not identify variants at any of the 1,919 SNPs with reference genotypes in the Omni 2.5 data, for a specificity of 1 (Table 1). The average sensitivity of

the GEDi test, including data from the nine replicates of all four samples, was 0.979 ± 0.007, and the specificity was 1 ± 0.

We investigated the false-negative base calls in the GEDi data further and found that there were 7–11 discrepancies per sample identified between the GEDi and Omni 2.5 data (Table 2). In total, there were discrepancies detected at 23 positions that were predominantly related to the heterozygous versus homozygous state of the same identified base, with a different base identified at only one position: chr15:78397352. The NGS data showed that the Omni data at chr15:78397352 were incorrect because of a single base deletion adjacent to the interrogated base, which shifted the base analyzed by the single-base extension method used in the Omni arrays (Figure 2a). Indels were associated with four additional GEDi versus Omni discrepancies, and all but one of the remaining differences were attributable to low SNP quality scores in the SAMtools variant identification software (Table 2). At one position (chr4:6304087), the Omni data were incorrect (confirmed by Sanger sequencing), without any evident explanation. A small number of bases (7–10) were not called in the GEDi data across all replicates (no call; Figure 2c), and Omni 2.5 SNP calls were not obtained for 11–19 positions (Omni no value; Figure 2c). Bases were scored as “no call/match” if one or more replicates

**Table 1** Sensitivity and specificity calculations for GEDi versus Omni 2.5 single-nucleotide polymorphism (SNP) data

2 × 2 Contingency table		
	Omni + (SNP)	Omni – (REF)
GEDi + (SNP)	True positive	False positive
GEDi – (REF)	False negative	True negative
	Sensitivity = true positive/(true positive + false negative)	Specificity = true negative/(true negative + false positive)
<b>NA12878</b>		
	Omni + (SNP)	Omni – (REF)
GEDi + (SNP)	508 (503–510, SD: 2)	0 (0–0, SD: 0)
GEDi – (REF)	12 (8–20, SD: 4)	1933 (1919–1944, SD: 13)
	Sensitivity: 0.977 (0.967–0.981, SD: 0.004)	Specificity: 1 (1–1, SD: 0)
<b>OGI-281–608</b>		
	Omni + (SNP)	Omni – (REF)
GEDi + (SNP)	469 (462–472, SD: 3)	0 (0–0, SD: 0)
GEDi – (REF)	13 (10–20, SD: 3)	1944 (1944–1944, SD: 0)
	Sensitivity: 0.973 (0.959–0.979, SD: 0.007)	Specificity: 1 (1–1, SD: 0)
<b>OGI-132–357</b>		
	Omni + (SNP)	Omni – (REF)
GEDi + (SNP)	495 (492–497, SD: 1)	0 (0–0, SD: 0)
GEDi – (REF)	10 (8–13, SD: 1.4)	1919 (1919–1919, SD: 0)
	Sensitivity: 0.981 (0.974–0.984, SD: 0.003)	Specificity: 1 (1–1, SD: 0)
<b>OGI-307–717</b>		
	Omni + (SNP)	Omni – (REF)
GEDi + (SNP)	508 (506–511, SD: 1.716)	0 (0–0, SD: 0)
GEDi – (REF)	7 (4–9 SD: 2)	1917 (1917–1917, SD: 0)
	Sensitivity: 0.986 (0.983–0.992, SD: 0.003)	Specificity: 1 (1–1, SD: 0)

For each DNA sample, the number of positions at which variants (single-nucleotide polymorphisms (SNPs)) or the reference (REF) were detected by the Omni 2.5 SNP arrays and the genetic eye disease (GEDi) test are indicated. For the GEDi data, the ranges derived from the nine replicates for each sample tested are shown. The average sensitivity and specificity for each DNA sample are shown, with the ranges and SDs included in parentheses. The overall sensitivity and specificity reported in the text are the averages of these data for all four samples. A standard 2 × 2 contingency table with definitions of sensitivity and specificity is shown for reference.

**Table 2** GEDi versus Omni 2.5 discrepancies detected

	NA12878	OGI-132–357	OGI-281–608	OGI-307–717	Discrepancy
chr1:156146218	—	—	1/9 HOM/HET	—	1
chr1:213071341	—	—	—	1/9 HOM/HET	1
chr2:62052380	—	—	1/9 HOM/HET	—	1
chr2:166770120	—	3/8 HOM/HET	—	—	1
chr3:150645351	2/9 HOM/HET	—	—	—	1
chr3:193413502	1/5 HOM/HET	—	—	—	1
chr4:15982166	—	—	—	2/7 HOM/HET	1
chr5:178405941	2/9 HOM/HET	—	—	1/9 HOM/HET	1
chr6:42932200	1/7 HOM/HET	1/3 HOM/HET	—	1/7 HOM/HET	1
chr6:42932202	1/8 HOM/HET	1/7 HOM/HET	—	—	1
chr9:102861613	—	—	1/9 HOM/HET	—	1
chr9:139327064	—	9/9 HOM/HET	9/9 HOM/HET	9/9 HOM/HET	1
chr10:73461805	—	—	1/9 HOM/HET	—	1
chr10:85976966	9/9 HOM/HET	9/9 HOM/HET	—	—	1
chr16:1265600	—	1/7 HOM/HET	—	—	1
chr16:1574863	—	1/9 HOM/HET	—	—	1
chr16:57937788	—	—	1/9 HOM/HET	—	1
chr17:11835331	—	—	1/9 HOM/HET	—	1
chr3:63986047	9/9 HOM/HET	—	9/9 HOM/HET	—	2
chr4:6304087	—	—	—	9/9 HOM/HET	5
chr4:15981874	—	9/9 HOM/HET	—	—	4
chr4:15982166	9/9 HOM/HET	—	9/9 HOM/HET	—	4
chr11:76895772	—	9/9 HOM/HET	9/9 HOM/HET	—	2
chr15:78397352	9/9 DISCREP	9/9 DISCREP	9/9 DISCREP	9/9 DISCREP	3

Base positions of discrepancies detected in at least one genetic eye disease (GEDi) sequence replicate and the Omni 2.5 single-nucleotide polymorphism data for each sample analyzed are shown. The number of replicates with alternative results are indicated. Fewer than nine replicates are indicated for some positions at which base calls were not made in some replicates. Discrepancies located in the top portion of the table were due to low SNP quality scores in SAMtools; the bottom portion of the table contains discrepancies specific to the GEDi-versus-Omni data comparisons. The reasons identified for the discrepancies are indicated in the last column: 1 = low SNP quality score; 2 = heterozygous deletion of target base; 3 = homozygous deletion adjacent to target base; 4 = heterozygous insertion adjacent to target base; 5 = Omni incorrect, reason uncertain. Note that position chr4:15982166 is included in both halves of the table.

DISCREP, discrepancy; HET, heterozygous; HOM, homozygous; SNP, single-nucleotide polymorphism.

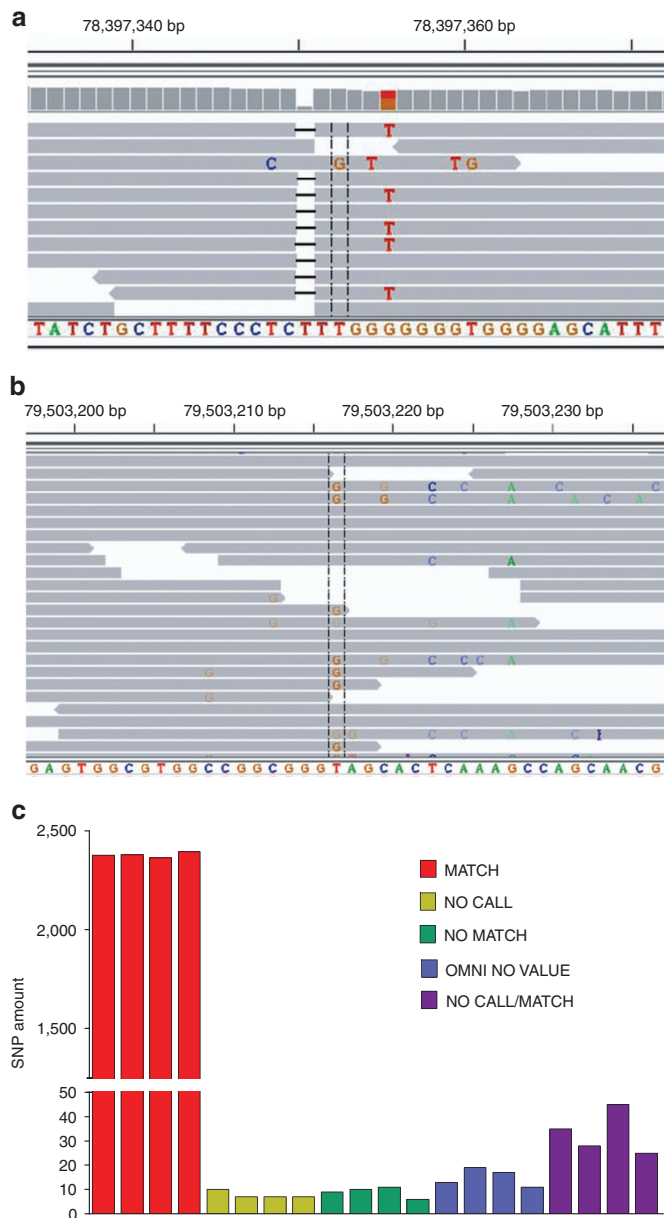
for each DNA sample had no call at that position but all other replicates matched. There were 25–45 of these bases, many of which (55/87 total = 63%) were due to no call in a single replicate (**Supplementary Table S2** online).

The accuracy of the GEDi test was also supported by comparison of the GEDi sequence data for the HapMap sample NA12878 with publically available Platinum 200× average depth WGS data for NA12878 from Illumina (<http://www.illumina.com/platinumgenomes>). Within the 1,197,667 bp in the GEDi capture regions, excluding the mitochondrial chromosome, there were 962 SNPs and 89 indels identified in the NA12878 WGS data by Illumina. The accuracy of the GEDi test in identifying both the SNPs and indels was 99.9%. The sensitivity and specificity for SNP detection were 96.4 and 99.9%, respectively, and for indel detection were 91.6 and 99.9%, respectively. It is likely that the sensitivity of the GEDi test for SNP detection is even higher; we identified 47 SNPs called in the Illumina Platinum data that are located in a highly repetitive 11-kb chr17 region (chr17:21311917–21323163) in the gene *KCNJ12* and that are likely to be incorrect because of poor read alignment (**Supplementary Figure S1** online).

### GEDi versus WES

The GEDi test performance was compared with WES by analyzing the WES data of the same four validation samples for the 2,443 Omni 2.5 SNPs in the GEDi gene set. The average DoC achieved by WES in these experiments was 100×, with 98% of the targeted regions covered at a 10× sequence depth. Using the Omni 2.5 data as the gold standard, the sensitivity of WES was  $0.883 \pm 0.004$  and the specificity of WES was  $0.9998 \pm 0.0003$ . Although both GEDi and WES have excellent specificity, this comparison shows that WES is ~10% less sensitive than the GEDi test. Analysis shows that this is because of a lack of sequence coverage in the WES data; ~10% of the 2,443 positions interrogated in these analyses have insufficient coverage ( $\geq 10\times$ ) to make an accurate base call (**Supplementary Table S3** online). The majority of these positions (76%) were common in all four samples, suggesting that these positions were covered less efficiently in the WES capture design. Comparison of the WES and GEDi capture baits at these positions confirmed this hypothesis and showed that 88% of the positions without coverage in the WES data had no baits in the V4+UTR capture set, whereas the GEDi capture set had at least one bait at these positions (**Supplementary Table S3** online).





**Figure 2 Sequencing accuracy.** (a) Integrative Genomics Viewer (IGV) screenshot of representative genetic eye disease (GEDi) next-generation sequencing (NGS) validation data at chr15:78397352. The Omni 2.5 single-nucleotide polymorphism (SNP) data were determined to be incorrect in all samples because of a single-base deletion adjacent to the interrogated base that shifted the analyzed base. (b) IGV screenshot of the putative c.1028T>G mutation of *FSCN2* in OGI-267-573, clarifying that the false-positive variant call was caused by misalignment of some NGS sequencing reads. (c) GEDI versus Omni 2.5 concordance histogram plot corresponding to the 2,443 shared SNPs between the GEDI design and Omni 2.5 SNP analysis of all 36 replicates of the four validation samples used in this study. MATCH, all GEDI NGS replicates matched Omni 2.5 SNP data; NO CALL, no NGS result; NO MATCH,  $\geq 1$  NGS replicate did not match Omni 2.5 SNP data; OMNI NO VALUE, no Omni 2.5 SNP result; NO CALL/MATCH,  $\geq 1$  NGS replicate had no result; all other NGS replicates matched Omni 2.5 SNP data.

An example of one of these regions is shown in **Figure 3**. Furthermore, there are 947 mutations in IRD genes reported in HGMD, ClinVar, and Ensemble that would be detected by

GEDi sequencing but that occur in regions that are not covered in the Agilent V4+UTR WES capture set (**Supplementary Table S4** online). Conversely, WES did detect bases at an average of 5.5 of 2,443 positions (0.22%) for which GEDI sequencing provided no call.

**Reproducibility and repeatability**

The reproducibility of variant detection by GEDI was assessed by comparing the detection rates for the 2,443 common SNPs in all 9 GEDI data sets from each of the 4 samples. Bases that were discrepant in one or more of the nine data sets for each sample in this GEDI-versus-GEDI comparison were identified. GEDI capture followed by Illumina sequencing is highly reproducible, with only four to six discrepancies detected between the replicates for each DNA sample. In each case the discrepancies were related to the heterozygous versus homozygous state of the same identified base. In the majority of cases, one of the nine sequence runs performed for each DNA sample contained the discrepancy (**Supplementary Table S5** online). Further analysis of the data for each of the 17 total discrepancies showed that they were due to a low SNP quality score in the SAMtools variant identification software. Sixteen of these are the same as those detected in the GEDI-versus-Omni 2.5 comparisons described above (**Supplementary Table S5** online).

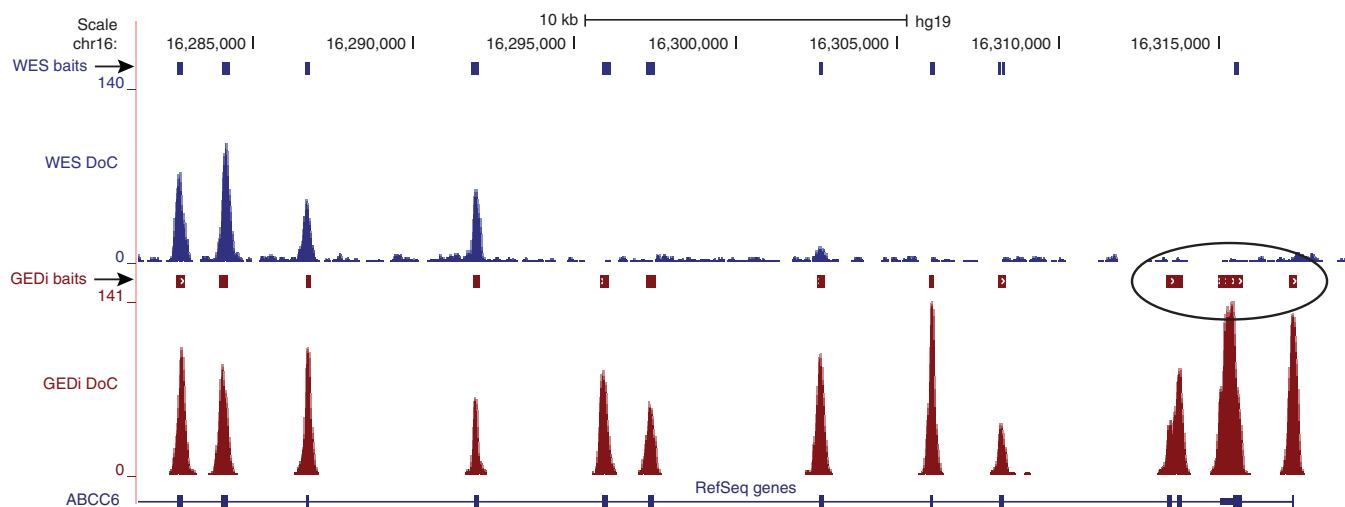
The repeatability and reproducibility of the GEDI test were also evaluated using the  $\kappa$  statistic, or  $\kappa$  coefficient of agreement.<sup>32</sup> For GEDI replicates performed on the same day, the  $\kappa$  statistic was 0.83088, indicating almost perfect agreement between the data obtained in the three replicates of each DNA sample analyzed.<sup>32</sup> For GEDI tests performed for the four individual samples on each of the three separate days, the  $\kappa$  statistic was 0.76366, indicating excellent reproducibility.<sup>32</sup>

**Mutation detection**

The GEDI test correctly identified mutations in 17 of 18 patient samples with known IRD, glaucoma, and optic atrophy variants, including 10 indels (**Supplementary Table S6** online). GEDI testing did not correctly identify the pathogenic *FOXC1* indel mutation in a patient with glaucoma; however, analysis showed a design gap in *FOXC1* where this pathogenic mutation is located.

**Clinical sensitivity**

GEDi clinical sensitivity was analyzed using samples from 192 probands with diagnoses of isolated or syndromic IRD, albinism, or microphthalmos (**Supplementary Table S7** online). Analyses of the sequence data identified genetic diagnoses for 98 of the probands, representing a clinical sensitivity of 51%, consistent with findings from other studies<sup>22-26</sup> (**Supplementary Table S7** online). The majority of these diagnoses were consistent with the subject’s clinical presentation and family history. Two subjects without a family history of disease had mutations in the known dominant IRD genes *PRPH2* and *IMPDH1*, consistent with identification of de novo mutations in the affected individuals; segregation analyses confirmed



**Figure 3 Comparison of V4+UTR whole-exome sequencing (WES) and genetic eye disease (GEDi) capture baits at the 5' end of *ABCC6*.** Read depth for the *ABCC6* gene is depicted for WES (top, blue) and GEDi sequencing (bottom, red). The baits contained in the WES and GEDi capture systems are indicated by boxes at the top of each section. As highlighted in the oval, the GEDi capture system has baits for all *ABCC6* exons, whereas some exons are not covered by baits in the V4+UTR exome capture system. The *ABCC6* gene model reference used is a “collapsed” reference that accounts for all known gene isoforms. DoC, depth of coverage; UTR, untranslated region.

de novo mutations in these two subjects (OGI-301–703 and OGI-274–582; **Supplementary Table S7** online). Although de novo mutations have been reported as the cause of dominant retinitis pigmentosa, de novo mutations in the *PRPH2* and *IMPDH1* genes have not been previously reported. The majority of subjects who were not diagnosed genetically by GEDi testing had nonsyndromic retinitis pigmentosa (52/89 = 58%; **Supplementary Table S8** online).

#### Mutation validation

We identified a total of 147 likely pathogenic mutations by GEDi capture and NGS sequencing, and all but 4 of these were validated by polymerase chain reaction and Sanger sequencing (**Supplementary Methods** online and **Supplementary Table S9** online). Review of the NGS data for the four putative mutations that were not validated by Sanger sequencing showed that three of the four mutations, corresponding to two probands (OGI-040–100 and OGI-271–579), were detected by fewer than 10 reads (**Supplementary Table S9** online). The fourth putative mutation not detected by Sanger sequencing had excellent DoC (**Supplementary Table S9** online); however, the heterozygous G base call was due to misalignment of some of the NGS sequence reads, resulting in a false-positive variant call (**Figure 2b**).

#### Missed diagnoses

GEDi capture and sequencing did not initially identify a genetic cause of disease in five patients for whom genetic diagnoses were ultimately obtained (**Supplementary Table S10** online). These cases are instructive, and information from them has been used to iteratively improve the GEDi test. For example, in two cases (OGI-147–394 and OGI-387–839), GEDi sequencing identified a single potentially pathogenic variant in *ABCA4* and *USH2A*, respectively, but the second mutant allele was

not initially detected (**Supplementary Table S10** online). The second alleles were subsequently identified by Sanger sequencing, both being deep intronic mutations known to alter splicing.<sup>31,33</sup> Probes for the relevant intronic region for *USH2A* have been added to subsequent versions of the GEDi capture set (**Supplementary Table S1b** online), and those corresponding to deep intronic *ABCA4* mutations will be added to the next version of GEDi.<sup>33</sup> Information regarding the remaining three cases listed in **Supplementary Table S10** online is included in the **Supplementary Material** online.

#### Improved diagnoses

Of note, seven of the subjects studied had mutations in genes that are not primarily associated with their phenotypes (denoted with an asterisk in **Supplementary Table S7** online). Specific examples include cone dystrophy due to mutations in the ORF15 region of *RPGR*, *NRL* mutations in chorioretinal atrophy, and *TMEM67* mutations in Senior–Loken syndrome (**Supplementary Figure S2** online; additional details are provided in the **Supplementary Material** online).

## DISCUSSION

Our results suggest that selective targeted enrichment with NGS is the preferred method for diagnostic testing, especially for genetically heterogeneous disorders such as IRDs. The GEDi test has improved sensitivity as compared with WES while maintaining nearly perfect specificity. Our results show that the higher sensitivity of the GEDi test is because of improved probe design as compared with that of a commercially available exome capture set in which probes were missing for ~10% of the regions targeted by the GEDi test. Although the concept that targeted sequencing can outperform standard exome sequencing based on better coverage has been discussed in reviews and commentaries regarding genetic diagnostic testing, only limited empiric

comparisons of these two approaches to genetic diagnostic testing have been reported previously.<sup>3,10,34</sup> Thus, although WES is now available as a clinical diagnostic test at some centers and reports of using WES for diagnostic testing have been published, quantitation of the performance characteristics of the GEDi test makes identifying and quantifying the advantages of selective targeted enrichment over WES possible.<sup>6,35,36</sup>

There are additional advantages of selective enrichment or panel tests over WES for diagnostic testing. The turnaround time for the GEDi “panel” test run on a MiSeq NGS platform is ~1 day, which is considerably less than that for WES samples run on a HiSeq 2000 instrument (~12 days). The current costs of selective exon capture tests are also lower than those for WES, although it is likely that this difference will continue to diminish over time. As of August 2014, the cost of the materials needed for GEDi testing per patient was ~\$430, as compared with \$1,325 per patient for WES using the sequence depth described. In addition, panel testing has a higher pretest probability of finding a meaningful result and reduces the potential for incidental sequence findings, which can be challenging for both health-care providers and patients.<sup>37,38</sup>

Although multiple characteristics make selective targeted enrichment a preferable method for genetic diagnostic testing, there are some drawbacks to this approach. First, hybridization-based capture approaches are limited by “design gaps,” regions where it is not possible to design targeted enrichment probes. Specifically, genome regions with high GC content and/or repetitive elements can be resistant to accurate capture probe design.<sup>3</sup> Fortunately, based on data obtained for the GEDi test, near-target sequence coverage limits this problem to gaps larger than 75 bp, which reduces, although does not eliminate, this problem. Using alternative approaches to capture regions in hybridization design gaps is also possible, including amplification-based strategies such as Agilent Technologies’ HaloPlex technique.<sup>39</sup>

Sufficient sequence depth is also needed to make accurate base calls from the NGS data. For GEDi, we showed that a minimum DoC of 10× gave a specificity of 100%. Even with sufficient DoC, misalignment of short NGS reads can lead to incorrect base identification, especially for repetitive regions or genes with paralogous copies elsewhere in the genome, which we observed in both the GEDi and WES data. Until longer sequencing reads become routinely available, this is likely to remain a problem; however, familial segregation studies and Sanger validation of pathogenic alleles can be useful in resolving these discrepancies.

We carefully evaluated the overall performance characteristics of the GEDi test and showed that the test is both sensitive and specific, as well as highly reproducible and accurate. Thorough analyses of these test characteristics have been reported for one other NGS-based diagnostic test called the WUCaMP assay, a targeted enrichment and NGS-based test for 25 genes associated with cancer.<sup>5</sup> The sensitivity and specificity of the WUCaMP assay were determined by comparing test data with WGS data from Complete Genomics for HapMap sample NA19240. For these studies, the test samples were sequenced to

a high DoC, with 96.9% of the targeted regions covered at ≥50× depth. The reported sensitivity and specificity for detecting single-nucleotide variations were 98.3 and 100%, respectively.<sup>5</sup> The sensitivity and specificity of the GEDi test for detecting single-nucleotide variations are comparable, at 96.4–97.9% and 99.9–100%, respectively, with an overall accuracy for both the SNPs and indels of 99.9%. We also showed that the GEDi test is highly repeatable and reproducible, with  $\kappa$  statistics of 0.83088 and 0.76366, respectively, indicating excellent agreement between the data obtained in the replicate testing of the four individual DNA samples.<sup>32</sup>

The clinical sensitivity of the GEDi test was 51% in patients with IRDs, a rate that is consistent with prior reports.<sup>22–26</sup> It is hypothesized that subjects without mutations in GEDi target genes must have mutations in novel disease genes or in non-coding portions of the currently identified IRD genes. Exome and genome sequencing will be required to identify these mutations. It is also possible that some subjects have mutations that cannot be readily detected by sequencing-based approaches, such as copy-number variations.<sup>25,40</sup>

Comprehensive genetic diagnostic testing for genetically and phenotypically heterogeneous disorders such as IRDs can also lead to diagnoses outside the reported genotype–phenotype relationships. Seven of the patients with genetic diagnoses had atypical phenotypic features, confirming that it can be difficult to predict the genetic cause of disease based on clinical findings alone.<sup>26</sup>

In summary, the GEDi test offers a number of advantages as a clinical diagnostic test for patients with inherited eye disorders. Given the potential for gene-based therapies for inherited disorders in general and inherited eye disorders in particular, genetic diagnostic testing will increasingly be necessary for the optimal care of patients with genetic diseases. Furthermore, the GEDi test statistics make a strong case for the use of targeted tests in the clinical setting because they are highly accurate and reproducible, and have better overall performance than more general tests such as conventional WES analyses.

#### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

#### ACKNOWLEDGMENTS

The authors are grateful to Aliete E. Langsdorf for the technical contribution and to Kendrick Goss for assistance with the PPT1 assays. This work was supported by grants from the National Institutes of Health (EY012910 (E.A.P.) and P30EY014104 (Massachusetts Eye and Ear Infirmary core support)), the March of Dimes (J.L.W.), and the Foundation Fighting Blindness USA (E.A.P. and Q.L.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding organizations or the National Institutes of Health. GEDi sequence data, WES sequence data, and Omni 2.5 SNP data for the four samples used to quantify the performance of the GEDi test are available at dbGAP (study accession no. phs000798.v1.p1). Also, the variants



listed in **Supplementary Table S7** online have been submitted to the ClinVar database.

## DISCLOSURE

The authors declare no conflict of interest.

## REFERENCES

- Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* 2012;4:118ra110.
- Sommen M, Van Camp G. Genetic diagnostics of early childhood hearing loss: better testing with next-generation DNA sequencing. *B-ENT* 2013;Suppl 21:51–56.
- Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet* 2013;14:295–300.
- Pugh TJ, Kelly MA, Gowrisankar S, et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med* 2014;16:601–608.
- Cottrell CE, Al-Kateb H, Bredemeyer AJ, et al. Validation of a next-generation sequencing assay for clinical molecular oncology. *J Mol Diagn* 2014;16:89–105.
- Levenson D. Whole-exome sequencing emerges as clinical diagnostic tool: testing method proves useful for diagnosing wide range of genetic disorders. *Am J Med Genet A* 2014;164A:ix–ix.
- Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 2012;30:1033–1036.
- Rehm HL, Bale SJ, Bayrak-Toydemir P, et al.; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;15:733–747.
- Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013;369:1502–1511.
- Shearer AE, Black-Ziegelbein EA, Hildebrand MS, et al. Advancing genetic testing for deafness with genomic technology. *J Med Genet* 2013;50:627–634.
- Buch H, Vinding T, La Cour M, Appleyard M, Jensen GB, Nielsen NV. Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults: the Copenhagen City Eye Study. *Ophthalmology* 2004;111:53–61.
- Congdon N, O'Colmain B, Klaver CC, et al.; Eye Diseases Prevalence Research Group. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol* 2004;122:477–485.
- Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006;90:262–267.
- Fan BJ, Wiggs JL. Glaucoma: genes, phenotypes, and new directions for therapy. *J Clin Invest* 2010;120:3064–3072.
- Gemenetzi M, Yang Y, Lotery AJ. Current concepts on primary open-angle glaucoma genetics: a contribution to disease pathophysiology and future treatment. *Eye (Lond)* 2012;26:355–369.
- Neuhann T, Rautenstrauss B. Genetic and phenotypic variability of optic neuropathies. *Expert Rev Neurother* 2013;13:357–367.
- Maguire AM, Simonelli F, Pierce EA, et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* 2008;358:2240–2248.
- Bainbridge JW, Smith AJ, Barker SS, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* 2008;358:2231–2239.
- Cideciyan AV, Aleman TS, Boye SL, et al. Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc Natl Acad Sci USA* 2008;105:15112–15117.
- MacLaren RE, Groppe M, Barnard AR, et al. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *Lancet* 2014;383:1129–1137.
- Wiggs JL, Pierce EA. Genetic testing for inherited eye disease: who benefits? *JAMA Ophthalmol* 2013;131:1265–1266.
- Song J, Smaoui N, Ayyagari R, et al. High-throughput retina-array for screening 93 genes involved in inherited retinal dystrophy. *Invest Ophthalmol Vis Sci* 2011;52:9053–9060.
- Audo I, Bujakowska KM, Léveillard T, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis* 2012;7:8.
- Neveling K, Collin RW, Gilissen C, et al. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* 2012;33:963–972.
- Eisenberger T, Neuhaus C, Khan AO, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One* 2013;8:e78496.
- Wang F, Wang H, Tuan HF, et al. Next generation sequencing-based molecular diagnosis of retinitis pigmentosa: identification of a novel genotype-phenotype correlation and clinical refinements. *Hum Genet* 2014;133:331–345.
- Falk MJ, Pierce EA, Consugar M, et al. Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discov Med* 2012;14:389–399.
- Falk MJ, Zhang Q, Nakamaru-Ogiso E, et al. NMNAT1 mutations cause Leber congenital amaurosis. *Nat Genet* 2012;44:1040–1045.
- den Hollander AI, Koenekoop RK, Yzer S, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet* 2006;79:556–561.
- Webb TR, Parfitt DA, Gardner JC, et al. Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Hum Mol Genet* 2012;21:3647–3654.
- Vaché C, Besnard T, le Berre P, et al. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Hum Mutat* 2012;33:104–108.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363.
- Braun TA, Mullins RF, Wagner AH, et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* 2013;22:5136–5145.
- Redin C, Le Gras S, Mhamdi O, et al. Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes. *J Med Genet* 2012;49:502–512.
- Flintoft L. Clinical genetics: exomes in the clinic. *Nat Rev Genet* 2013;14:824.
- Delanty N, Goldstein DB. Diagnostic exome sequencing: a new paradigm in neurology. *Neuron* 2013;80:841–843.
- Dorschner MO, Amendola LM, Turner EH, et al.; National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 2013;93:631–640.
- Platt J, Cox R, Enns GM. Points to consider in the clinical use of NGS panels for mitochondrial disease: an analysis of gene inclusion and consent forms. *J Genet Couns* 2014;23:594–603.
- Berglund EC, Lindqvist CM, Hayat S, et al. Accurate detection of subclonal single nucleotide variants in whole genome amplified and pooled cancer samples using HaloPlex target enrichment. *BMC Genomics* 2013;14:856.
- Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* 2012;91:597–607.