Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing

A. Eliot Shearer[a,b,1], Adam P. DeLuca[c,d,1], Michael S. Hildebrand[a,1], Kyle R. Taylor[c,d], José Gurrola II[a], Steve Scherer[e,2], Todd E. Scheetz[c,d,f,2], and Richard J. H. Smith[a,b,g,2,3]

[a]Department of Otolaryngology, Head and Neck Surgery, University of Iowa, Iowa City, IA 52242; [b]Department of Molecular Physiology and Biophysics, University of Iowa Carver College of Medicine, Iowa City, IA 52242; [c]Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242; [d]Center for Bioinformatics and Computational Biology, University of Iowa, Iowa City, IA 52242; [e]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; [f]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242; and [g]Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA 52242

Edited* by Mary-Claire King, University of Washington, Seattle, WA, and approved October 14, 2010 (received for review August 31, 2010)

The extreme genetic heterogeneity of nonsyndromic hearing loss (NSHL) makes genetic diagnosis expensive and time consuming using available methods. To assess the feasibility of target-enrichment and massively parallel sequencing technologies to interrogate all exons of all genes implicated in NSHL, we tested nine patients diagnosed with hearing loss. Solid-phase (NimbleGen) or solution-based (SureSelect) sequence capture, followed by 454 or Illumina sequencing, respectively, were compared. Sequencing reads were mapped using GSMAPPER, BFAST, and BOWTIE, and pathogenic variants were identified using a custom-variant calling and annotation pipeline (ASAP) that incorporates publicly available in silico pathogenicity prediction tools (SIFT, BLOSUM, Polyphen2, and Align-GVGD). Samples included one negative control, three positive controls (one biological replicate), and six unknowns (10 samples total), in which we genotyped 605 single nucleotide polymorphisms (SNPs) by Sanger sequencing to measure sensitivity and specificity for SureSelect-Illumina and NimbleGen-454 methods at saturating sequence coverage. Causative mutations were identified in the positive controls but not in the negative control. In five of six idiopathic hearing loss patients we identified the pathogenic mutation. Massively parallel sequencing technologies provide sensitivity, specificity, and reproducibility at levels sufficient to perform genetic diagnosis of hearing loss.

deafness | genomics | Usher syndrome | diagnostics | next-generation sequencing

Hereditary sensorineural hearing loss (SNHL) is the most common sensory impairment in humans (1, 2). In developed countries, two-thirds of prelingual-onset SNHL is estimated to have a genetic etiology, of which ∼70% is nonsyndromic hearing loss (NSHL). Eighty percent of NSHL is autosomal recessive nonsyndromic hearing loss (ARNSHL), ∼20% is autosomal dominant (AD), and the remainder is composed of X-linked and mitochondrial forms (1, 3). To date, 134 deafness loci have been identified, and 32 recessive (DFNB), 23 dominant (DFNA) and 2 X-linked (DFNX) genes have been cloned; 8 genes are associated with both ARNSHL and ADNSHL (4).

Establishing a genetic diagnosis of NSHL is a critical component of the clinical evaluation of deaf and hard-of-hearing persons and their families. If a genetic cause of hearing loss is determined, it is possible to provide families with prognostic information, recurrence risks, and improved habilitation options. For persons diagnosed with Usher syndrome, preventative measures including sunlight protection and vitamin therapy can be implemented to minimize the rate of progression of retinitis pigmentosa (5). Most current genetic testing strategies for NSHL rely on a gene-specific Sanger sequencing approach. Because mutations in a single gene, *GJB2* (DFNB1), account for up to 50% of ARNSHL in many world populations (6), this approach has changed the evaluation of patients with presumed ARNSHL. However, the mutation frequency in other genes in persons with NSHL in outbred populations is unknown, making sequential gene screening problematic (7). The extreme heterogeneity of NSHL also makes serial sequencing approaches unfavorable in terms of efficiency and cost.

The advent of new technologies that target and enrich specific regions of the genome coupled with massively parallel sequencing offers an alternative approach to genetic testing for deafness. Although genetic diagnoses can also be made by whole genome sequencing (8) or targeted sequence capture of the entire exome (9, 10), these approaches are expensive, and time-consuming data analysis is required. Therefore, our aim was to develop and test a streamlined, comprehensive genetic diagnostic platform that targets only the 0.014% of the genome currently associated with NSHL.

Results

We elected to compare the two most widely used target enrichment approaches (NimbleGen solid-phase and SureSelect solution-based sequence capture) and two massively parallel sequencing technologies (454 GS FLX pyrosequencing and Illumina GA_{II} cyclic reversible termination sequencing) using genomic DNA from NSHL families. The optimized platform and diagnostic pipeline, which we refer to as OtoSCOPE (otologic sequence capture of pathogenic exons), targets the exons of all 54 known deafness genes (Table S1). The subjects were nine individuals with presumed NSHL (Table 1). Sanger sequencing-based genetic testing had been completed for *GJB2* and *SLC26A4* in all autosomal recessive (AR) cases. Positive controls had been genetically diagnosed by Sanger sequencing prior to this study.

We performed array and solution-based targeted capture of all exons of the 54 genes known to cause NSHL, also including Usher syndrome genes because in infants and children, Usher syndrome is not readily distinguishable from ARNSHL (*Materials and Methods* and Table S1). Some requested target regions were not covered by the NimbleGen and SureSelect designs (9.3% and 8.3%, respectively; Table 2) due to repetitive regions. These percentages are consistent with other reports focused on exon sequence capture (11, 12). However, the proportion of protein coding regions included in the SureSelect bait design was comparatively better than that covered by NimbleGen (97.7 and 93.6%, respectively; Table 2), reflecting different methods for defining repetitive regions and complementary oligonucleotide selection.

Author contributions: A.E.S., M.S.H., S.S., T.E.S., and R.J.H.S. designed research; A.E.S., A.P.D., M.S.H., and K.R.T. performed research; A.P.D., K.R.T., J.G., S.S., and T.E.S. contributed new reagents/analytic tools; A.E.S., A.P.D., M.S.H., K.R.T., J.G., T.E.S., and R.J.H.S. analyzed data; and A.E.S., M.S.H., and R.J.H.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]A.E.S., A.P.D., and M.S.H. contributed equally to this work.

[2]S.S., T.E.S., and R.J.H.S. contributed equally to this work.

[3]To whom correspondence should be addressed. E-mail: richard-smith@uiowa.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012989107/-/DCSupplemental.

Downloaded at Palestinian Territory, occupied on December 29, 2021

www.manaraa.com

Downloaded at Palestinian Territory, occupied on December 29, 2021

Table 1. Samples studied and diagnostic sequencing results by allele

Sample	Status	Hearing loss inheritance	Candidate variants	Gene	Nucleotide change	Protein change	Genomic coordinates (Hg19)	Sequencing reads–variant reads/total coverage (%)	Status
1	Positive control	ADNSHL	3	*COCH*	c.151C > T	p.P51S*	Chr14: 31,346,846	463/902 (51.3)	Pos Ctrl mutation
				EYA4	c.390C > A	p.Tyr130X	Chr6: 133,782,271	234/673 (34.77)	RO-segregation
				OTOF	c.4484G > C	p.R1495P	Chr2: 26,689,598	639/1,266 (50.5)	RO-segregation
2	Positive control	ARNSHL	1	*GJB2*	c.109G > A	p.V37I†	Chr13: 20,763,612	1,075/1,078 (99.7)	Pos Ctrl mutation
				GJB2	c.109G > A	p.V37I	Chr13: 20,763,612	1,075/1,078 (99.7)	Pos Ctrl mutation
3	Replicate of 2	ARNSHL	1	*GJB2*	c.109G > A	p.V37I	Chr13: 20,763,612	1,113/1,114 (99.9)	Pos Ctrl mutation
				GJB2	c.109G > A	p.V37I	Chr13: 20,763,612	1,113/1,114 (99.9)	Pos Ctrl mutation
4	Negative control	n/a	0	—	—	—	—	—	—
5	Idiopathic SNHL	ARNSHL	1	*STRC*	c.4057C > T	p.Q1353X	Chr15: 43,896,918	44/47 (93.6)	Causative mutation
				STRC	~100-kb deletion	—	Chr15: q15	—¶	Causative mutation
6	Idiopathic SNHL	ADNSHL	2	*MYO6*	c.862_865delACAA	p.D288DfsX17	Chr6: 76,554,660–76,554,663	109/306 (35.6)	Causative mutation
				MYO6	c.3667G > A	p.D1223N	Chr6: 76,624,538	427/890 (48.0)	RO-segregation
7	Idiopathic SNHL	ADNSHL	4	*MYO7A*	c.5156A > G	p.Y1719C	Chr11: 76,913,457	439/908 (48.3)	RO-segregation
				PCDH15	c.1039C > T	p.L347F	Chr10: 55,973,755	215/459 (46.8)	RO-segregation
				USH2A	c.6713A > C	p.E2238A	Chr1: 216,166,454	372/752 (49.47)	RO-segregation
				OTOF	c.3751T > G	p.C1251G	Chr2: 26,695,500	706/720 (98.1)	RO-segregation
8	Idiopathic SNHL	ADNSHL	2	*KCNQ4*	c.842T > C	p.L281S‡	Chr1: 41,285,554	348/893 (39.0)	Causative mutation
				WFS1	c.2327A > T	p.E776V	Chr4: 6,303,849	161/321 (50.1)	RO-segregation
9	Idiopathic SNHL	ADNSHL	2	*MYH14*	c.5893G > T	p.E1965X	Chr19: 50,812,952	587/1,255 (46.8)	Causative mutation
				MYH14	c.4193A > G	p.E1398G	Chr19: 50,784,876	413/1,334 (31.0)	RO-controls
10	Idiopathic SNHL	ARNSHL	2	*CDH23*	c.1096G > A	p.A366T§	Chr10: 73,377,112	2,233/4,529 (49.3)	Causative mutation
				CDH23	c.0.3293A > G	p.N1098S§	Chr10: 73,472,494	2,708/5,445 (49.7)	Causative mutation

Candidate variants follow inheritance pattern (if known) and are predicted deleterious by Polyphen2, SIFT, BLOSUM62, and align-GVGD. Candidate mutations were ruled out or in by segregation analysis and/or presence in controls. Causative mutations are predicted damaging, segregate with hearing loss in family, are not found in controls, and/or are known to cause hearing loss. NSHL, nonsyndromic hearing loss; ADNSHL, autosomal dominant nonsyndromic hearing loss; ARNSHL, autosomal recessive nonsyndromic hearing loss; Pos Ctrl, positive control; RO-segregation, mutation ruled out by segregation analysis; RO-controls, mutation ruled out by controls analysis.

*Previously reported (16).

†Previously reported (17).

‡Previously reported (18).

§Previously reported (19).

¶Deletion determined by sequencing reads analysis, confirmed by PCR and array-CGH, previously reported (ref. 20 and Fig. 1).

www.manaraa.com

Table 2. Sequence capture performance results

	NimbleGen-454		SureSelect-Illumina	
	All regions (%)	Protein coding regions (%)	All regions (%)	Protein coding regions
Number of bases requested*	421,741 bp	187,017 bp	421,741 bp	187,017 bp
Bases covered by complementary oligonucleotides (%)	386,880 bp (91.7)	266,947 bp (93.6)	382,627 bp (90.7)	182,749 bp (97.7)
Mean sequence coverage in bases (%)	406,756 bp (96.4)	180,489 bp (96.5)	401,252 bp (95.1)	182,727 bp (97.7)
Mean sequence coverage at variant calling threshold in bases (%)†	403,873 bp (95.8)	179,468 bp (96.0)	385,101 bp (91.3)	178,230 bp (95.3)

Values given are a mean of all samples for each method.
*The same bases were requested for targeting using either NimbleGen or SureSelect and default repetitive regions were avoided in both methods.
†Variant calling threshold: for Illumina-SureSelect, 40× coverage and >30% of reads; for 454, ≥3× reads (multidirectional) and >30% of reads.

We paired NimbleGen solid phase sequence capture with 454 GS FLX pyrosequencing (NimbleGen-454 method) and SureSelect solution-based sequence capture with Illumina GAII sequencing (SureSelect-Illumina method) to determine the efficacy of these approaches for clinical diagnostics. Samples 1 and 2 were sequenced using both the NimbleGen-454 and SureSelect-Illumina to provide direct comparison of these methods, whereas samples 3–10 were examined only using SureSelect-Illumina.

The results of our study show that due to the disparity in sequence output between the two sequencing platforms (13), the overall sequence depth of coverage was on average 13-fold higher for the SureSelect-Illumina method as compared with the NimbleGen-454 method (903× and 71× depth of coverage, respectively (*Materials and Methods*); Table 3). However the percent of on-target reads (i.e., capture efficiency) was significantly lower ($P < 0.0001$) for the SureSelect-Illumina method (average 19.6%) than the NimbleGen-454 method (average 64.2%), which may reflect PCR-induced bias during Illumina library preparation, small target size, and sequencer oversaturation (Tables 2 and 3).

As a threshold for variant detection, we required ≥3 reads (multidirectional) for 454 and ≥40 reads for Illumina, with the added stipulation that the variant had to be present in at least 30% of reads to be considered high quality. Similar thresholds have been used in other studies (14) and reflect the need for high sensitivity in diagnostic testing (false positives are preferred over false negatives) (15). At these thresholds, 96.0% and 95.3% of protein coding bases were covered by the NimbleGen-454 and SureSelect-Illumina platforms, respectively (Table 2).

To validate the platforms as diagnostic tests we genotyped 605 highly heterozygous SNPs (average heterozygosity, 0.46; average 55 SNPs per sample) in the targeted regions by Sanger sequencing. Homozygous-alternative allele calls (n = 149; average 14 per sample) and heterozygous allele calls (n = 199; average 18 per sample) were used as true positive variants; homozygous-reference allele calls (n = 257; average 23 per sample) were used as true negative variants (simulated nonvariants). Variants not covered at variant-calling threshold for either platform were considered false negatives.

With both the NimbleGen-454 and SureSelect-Illumina methods, we identified the causative mutations in samples 1 and 2. Specificity was ∼98% for NimbleGen-454 and >99% for SureSelect-Illumina (Table 3). For the SureSelect-Illumina method, one false negative was found across all 10 samples (sensitivity 99.72%) reflecting high coverage depth of targeted regions. For the NimbleGen-454 method, a mean depth of coverage of 71× was associated with an increased false-negative rate (average two per sample) and a decrease in sensitivity (93.98%). This difference may be attributable to a relatively lower quality depth-of-coverage threshold for the NimbleGen-454 method (3× versus 40× for SureSelect-Illumina) as mean sequence coverage for protein-coding sequence was similar.

On the basis of data from samples 1 and 2, we analyzed the remaining eight samples using the SureSelect-Illumina method and a sequence analysis protocol we developed (*Materials and Methods*). In total, we detected causative mutations in the three positive controls and in five of six persons with idiopathic hearing loss (Table 1). Analysis of controls (samples 1–4) was completed in a blinded fashion with only the mode of inheritance known. Sample 1 was from a family segregating ADNSHL. Of three candidate variants, the known hearing loss mutation in *COCH* (c.151C > T, p.P51S, rs28938175; DFNA9) was the only variant that segregated with the phenotype in this family (16). In samples 2 and 3 (biological replicates), the only candidate variant determined in each sample was the known deafness mutation in *GJB2* (c.109G > A, p.V37I; DFNB1) (17). In sample 4, the negative control, none of the nine nonsynonymous/splice site/indel variations identified was predicted to be pathogenic.

Of the six persons with idiopathic hearing loss, two were from families segregating ARNSHL. Sample 5 was from a person with profound SNHL who was shown to carry a variant of unknown significance (VUS) in *STRC* (c.4057C > T, p.Q1353×; rs2614824; DFNB16) in 44/47 Illumina reads. This variant was not observed

Table 3. Sequencing results

Sequencing method	Capture method	Total sequencing reads	Uniquely mapping reads	% of uniquely mapping reads	Uniquely mapping reads on target (<1KB)	Mean coverage of targeted regions	Sens/spec	FN/FP
454	NimbleGen solid-phase	303,552	293,368	96.7%	194,155	71 X	93.98%/97.92%	2/0
Illumina	Agilent solution-based	41,234,307	33,693,349	81.3%	6,941,755	903 X	99.72%/>99%	0*/0

Illumina sequencing was performed using one sample per flow cell channel; 454 sequencing was performed using one sample per quarter of a four-way gasket. Values given are the mean of all samples for a given method: Illumina, 8 samples; 454, 2 samples. Sens, sensitivity; spec, specificity; FN, false negatives; FP, false positives.
*One FN was found across all ten samples.

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

in 100 ethnically matched controls (200 chromosomes). A hemizygous deletion of ∼100 kb involving the *STRC*–ψ*STRC* region was also detected (Fig. 1) and independently verified by PCR and array-comparative genomic hybridization (CGH). Homozygosity for this contiguous gene deletion is known to cause autosomal recessive deafness–infertility syndrome (20), making this person a compound heterozygote for a novel point mutation in *STRC in trans* with a large contiguous gene deletion that includes *STRC* (DFNB16). Sample 10, from a 4-y-old cochlear implant recipient with no family history of hearing loss, was compound heterozygous for known USH1D mutations in *CDH23* (c.1096G > A, p.A366T; c.3293A > G, p.N1098S) (19).

The remaining four samples were from families segregating ADNSHL. For these samples, we also used, when possible, the computer program AudioGene (21) as a phenotypic filter to predict probable ADNSHL genotypes on the basis of audiometric criteria through audioprofiling. Sample 6 carried two candidate variants in *MYO6* (DFNA22). The missense mutation was ruled out by segregation analysis; however, the 4-bp deletion (c.862_865delACAA, p.D288DfsX17) segregated with the phenotype and was not found in 135 ethnically matched controls (270 chromosomes). Because the DFNA22 locus is not represented on AudioGene, phenotypic filtering was not applied; however, the phenotype was consistent with reported DFNA22 audioprofiles (22). Sample 8 carried two candidate variants—one mutation was ruled out by segregation analysis and the other is a known ADNSHL mutation [(DFNA2) c.842C > T, p.L281S] in *KCNQ4* (18). This finding was consistent with AudioGene analysis, which predicted DFNA2 as the most likely cause of hearing loss in this family. Sample 9 carried two candidate variants—both detected in *MYH14*—that segregated with the phenotype in the extended family. One mutation was ruled out by analysis of controls, whereas the other, a stop mutation (c.5893G > T, p.E1965×), was not found in 119 ethnically matched controls (238 chromosomes) and was consistent with the predicted DFNA4 audioprofile. Four candidate variants were detected in sample 7, all of which were ruled out by segregation analysis. AudioGene analysis predicted *KCNQ4 (*DFNA2) as the most likely cause of hearing loss in this family; however, no variants in *KCNQ4* were identified in this analysis or by prior Sanger sequencing. This finding suggests that this family may segregate a novel genetic cause of ADNSHL, which is possible because causative genes have not been identified for more than half of the mapped ADNSHL loci (4).

Discussion

In this study, we identified NSHL mutations, three of which have not been reported, in eight of nine persons tested. Our sequencing results are consistent with other studies reporting utility of these technologies for diagnosis of other genetic diseases (11, 15, 23) and suggest that massively parallel sequencing is suitable for genetic testing of NSHL. A key concern for any diagnostic test is sensitivity, as it is critical that pathogenic mutations are not missed. This requirement could be viewed as a potential limitation of target-enrichment methodologies, as a significant portion of targeted bases in repetitive regions cannot be captured (Table 2). However, this limitation must be weighed against the decreased cost and time in simultaneously sequencing a large number of genes.

Multiplexing samples offers the opportunity to increase throughput, but to maintain read depth, capture efficiency must be increased (24, 25). Scalability and cost also must be considered for large-scale sequencing projects. Our results show that when adequate target coverage and depth of coverage are maintained, both the SureSelect-Illumina and NimbleGen-454 platforms offer relatively high specificity and sensitivity. However, the SureSelect-Illumina method is superior in terms of scalability, cost, and increased sensitivity. Our results compare very favorably with recent reports of high-throughput diagnostic tests for NSHL that rely on primer extension arrays (26) or targeted resequencing arrays (27), which are appealing due to low cost but are limited in terms of capacity (not all NSHL genes can be screened simultaneously) and therefore sensitivity.

In summary, we have demonstrated that OtoSCOPE has the potential to improve the efficiency of genetic testing for NSHL and Usher syndrome. Our results show that targeted capture plus massively parallel sequencing has a sensitivity and specificity comparable to Sanger sequencing. Comprehensive genetic screening for deafness using platforms like OtoSCOPE would allow clinicians to improve patient care by providing prognostic information and genetic counseling, and in cases like Usher syndrome, offer families preventative strategies to minimize the rate of progression of retinitis pigmentosa. Some of the novel habilitation options under

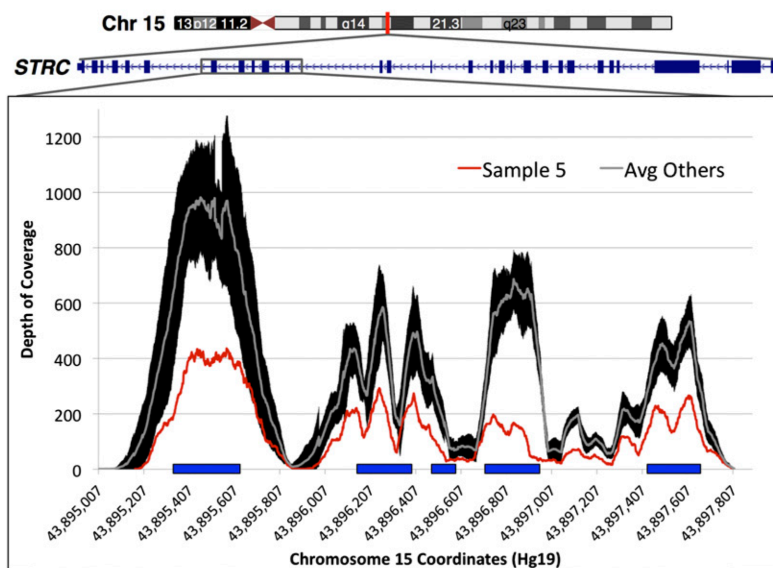

Fig. 1. Deletion analysis of sample 5 using massively parallel sequencing. Location on chromosome 15 is shown, with the highlighted region containing *STRC*; five exons are indicated by blue bars on *x* axis. Gray line is average sequencing depth of coverage for nine samples (all samples excluding sample 5); thick black line represents SD for these samples. Red line is depth of coverage for sample 5.

MEDICAL SCIENCES
Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

Table 4. Variant prioritization

Method	NimbleGen-454	SureSelect-Illumina
Variations in targeted genes at calling threshold*	1,008 (998–1,008)	876 (732–1112)
Variations with <1% or unknown allele frequency†	337 (335–339)	404 (297–507)
Synonymous variations	22 (17–26)	22 (18–25)
NS/SS/Indel variations	17 (15–18)	32 (21–40)
Individual-unique NS/SS/Indel variations‡	12 (9–14)	8 (3–14)
Predicted pathogenic variations	3 (2–4)	2 (0–3)
Candidate variants	2 (2)	2 (0–3)

Data given are means (range) (refer to Table S2 for detailed analysis). NS, nonsynonymous; SS, splice site.
*Variant calling threshold included only variants seen in 30% of reads and 40× Illumina depth of coverage or 3× multidirectional reads for 454.
†Allele frequency data taken from dbSNP130.
‡Individual unique variants excluded variants found in more than one sample in this study.

development to slow progression of hearing loss are also gene and even mutation specific (28), suggesting that comprehensive genetic testing will be an integral part of the care of deaf and hard-of-hearing patients in the future.

Materials and Methods

Patients. The subjects were nine individuals with presumed NSHL (Table 1) who provided informed consent for this study approved by the University of Iowa International Review Board. Detailed family histories, clinical evaluations, and audiograms were available for each patient.

Targeted Capture and DNA Sequencing. Genomic DNA (gDNA) was extracted from whole blood using standard procedures (29), quality assessed on an agarose gel and spectrophotometer, and quantified using the Qubit (Invitrogen) system. Three or 5 μg of gDNA were used for the SureSelect or NimbleGen capture methods, respectively.

Exons in all isoforms of the 54 known deafness genes were identified in the RefSeq and Ensembl databases using the University of California Santa Cruz table browser (http://genome.ucsc.edu/). An additional 50 bp of flanking intronic sequence were added to each exon and genomic intervals were merged using Galaxy software (http://galaxy.psu.edu). In total, we targeted 1,258 regions comprising 421,741 bp using both NimbleGen and SureSelect methods. The same genomic coordinates were sent to Roche for complementary oligonucleotide microarray design or uploaded to Agilent's eArray website for cRNA bait design.

DNA isolated using NimbleGen solid phase sequence capture was sequenced using 454 GS FLX pyrosequencing (454 Life Sciences); DNA obtained from SureSelect solution-based sequence capture was subjected to Illumina GA_{II} sequencing (Illumina). In both cases sequencing was performed according to manufacturer's protocols.

Sequence Analysis. Sequencing depth of coverage was defined as the number of sequencing reads, which had been filtered and mapped, per base. Average depth of coverage for each sequencing method was defined as the total depth of coverage per base, averaged over all requested bases. Depth of coverage per variant was defined as number of reads in which a variant was seen divided by the total depth of coverage for that base.

To prioritize a VUS, we developed a ranking algorithm that included the variant-calling threshold for each platform, location within the targeted region, type of change (nonsynonymous, splice-site, or frameshift deletion) and observed population frequency if the VUS was a reported SNP (Table 4). Other investigators have filtered variants on the basis of absence from the dbSNP database (9, 10, 14), a criterion we did not include as several mutations known to cause hearing loss have been assigned RefSNP (RS) numbers. Instead, we filtered variants on the basis of quantitative data from dbSNP130 such that any variant with an allele frequency >1% was considered a benign polymorphism (excluding known ARNSHL-associated variants of *GJB2*), whereas any variant with an allele frequency <1% (or of unknown frequency) was investigated further. In addition, we developed an in-house list of nonpathogenic nonsynonymous variants located within the targeted genes that can be excluded due to their presence in more than one nonreplicate sample or the negative control (list available upon request).

Pathogenicity of a nonsynonymous, splice-site or insertion-deletion (indel) VUS was assessed using in silico mutation prediction software. We incorporated four algorithms (BLOSUM62, SIFT, PolyPhen2, and Align-GVGD) that have the highest positive predictive value (94.6%) when concordant (30) and we required concurrence in at least three of four prediction tools. We also prioritized gene variants in the context of inheritance pattern (i.e., genes known to cause ARNSHL or ADNSHL) (Table 4 and Table S1). A phenotypic filter was applied to ADNSHL using a computer learning algorithm we have developed called AudioGene, which predicts probable genotypes on the basis of audiometric criteria by constructing audioprofiles (20). Candidate mutations were verified by Sanger sequencing in the extended families and if unreported, excluded in ethnically matched controls.

ACKNOWLEDGMENTS. We thank the following individuals at the Baylor College of Medicine's Human Genome Sequencing Center for their invaluable contributions to this work: Dr. Richard Gibbs (director), Donna Muzny (director of operations), Dr. Yi Han, Dr. Min Wang, and Fiona Ongeri. We also thank Dr. Benjamin Darbro (University of Iowa) for assistance with the array CGH. Funding was provided by National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders Grant R01 DC02842 (to R.J.H.S.), a National Health and Medical Research Council Overseas Biomedical Postdoctoral Training Fellowship (to M.S.H.), NIH Predoctoral Research Fellowship T32 GM082729 (to A.P.D.), and a Doris Duke Clinical Research Fellowship (to A.E.S.).

1. Smith RJ, Bale JF, Jr., White KR, White KR (2005) Sensorineural hearing loss in children. *Lancet* 365:879–890.
2. Morton NE (1991) Genetic epidemiology of hearing impairment. *Ann N Y Acad Sci* 630:16–31.
3. Van Camp G, Willems PJ, Smith RJ (1997) Nonsyndromic hearing impairment: Unparalleled heterogeneity. *Am J Hum Genet* 60:758–764.
4. Van Camp G, Smith RJH (2010) Hereditary Hearing Loss Homepage. Available at http://www.hereditaryhearingloss.org. Accessed June 8, 2010.
5. Hamel C (2006) Retinitis pigmentosa. *Orphanet J Rare Dis* 1:40.
6. Green GE, et al. (1999) Carrier rates in the midwestern United States for GJB2 mutations causing inherited deafness. *JAMA* 281:2211–2216.
7. Hilgert N, Smith RJ, Van Camp G (2009) Forty-six genes causing nonsyndromic hearing impairment: Which ones should be analyzed in DNA diagnostics? *Mutat Res* 681: 189–196.
8. Lupski JR, et al. (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362:1181–1191.
9. Ng S, et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35.
10. Choi M, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 106:19096–19101.
11. Tewhey R, et al. (2009) Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 10:R116.
12. Gnirke A, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–189.
13. Metzker ML (2010) Sequencing technologies: The next generation. *Nat Rev Genet* 11: 31–46.
14. Ng SB, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461:272–276.
15. Harismendy O, et al. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 10:R32.
16. Hildebrand MS, et al. (2009) Mutation in the COCH gene is associated with superior semicircular canal dehiscence. *Am J Med Genet A* 149A:280–285.
17. Abe S, Usami S, Shinkawa H, Kelley PM, Kimberling WJ (2000) Prevalent connexin 26 gene (GJB2) mutations in Japanese. *J Med Genet* 37:41–43.
18. Talebizadeh Z, Kelley PM, Askew JW, Beisel KW, Smith SD (1999) Novel mutation in the KCNQ4 gene in a large kindred with dominant progressive hearing loss. *Hum Mutat* 14:493–501.

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

19. Oshima A, et al. (2008) Mutation profile of the CDH23 gene in 56 probands with Usher syndrome type I. *Hum Mutat* 29:E37–E46.
20. Zhang Y, et al. (2007) Sensorineural deafness and male infertility: A contiguous gene deletion syndrome. *J Med Genet* 44:233–240.
21. Hildebrand MS, et al. (2008) Audioprofile-directed screening identifies novel mutations in KCNQ4 causing hearing loss at the DFNA2 locus. *Genet Med* 10:797–804.
22. Hilgert N, et al. (2008) A splice-site mutation and overexpression of MYO6 cause a similar phenotype in two families with autosomal dominant hearing loss. *Eur J Hum Genet* 16:593–602.
23. Morgan JE, et al. (2010) Genetic diagnosis of familial breast cancer using clonal sequencing. *Hum Mutat* 31:484–491.
24. Mamanova L, et al. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118.
25. Lee H, et al. (2009) Improving the efficiency of genomic loci capture using oligonucleotide arrays for high throughput resequencing. *BMC Genomics* 10:646.
26. Rodriguez-Paris J, et al. (2010) Genotyping with a 198 mutation arrayed primer extension array for hereditary hearing loss: Assessment of its diagnostic value for medical practice. *PLoS ONE* 5:e11804.
27. Kothiyal P, et al. (2010) High-throughput detection of mutations responsible for childhood hearing loss using resequencing microarrays. *BMC Biotechnol* 10:10.
28. Hildebrand MS, et al. (2008) Advances in molecular and cellular therapies for hearing loss. *Mol Ther* 16:224–236.
29. Grimberg J, et al. (1989) A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res* 17:8390.
30. Chan PA, et al. (2007) Interpreting missense variants: Comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat* 28:683–693.

Downloaded at Palestinian Territory, occupied on December 29, 2021

www.manaraa.com