DEAFNESS GENE DISCOVERY IN THE GENOMIC ERA

Tal Elkan-Miller* and Karen B. Avraham
Department of Human Molecular Genetics and Biochemistry,
Sackler Faculty of Medicine and Sagol School of Neuroscience,
Tel Aviv University, Tel Aviv, Israel

ABSTRACT

Hereditary deafness is a genetically heterogeneous disease, with wide variation in clinical symptoms displayed, including age of onset, frequencies affected, and changes in hearing threshold. As a result, there are some common and many more rare forms of deafness, each represented by mutations in different genes. The major challenge and goal in the field has been to identify the genes containing mutations leading to deafness, and to elucidate the auditory and vestibular pathways the proteins they encode are involved in. A remarkable amount of data has accumulated over the years to address the mechanisms associated with normal hearing and pathology of the inner ear. To date, there are 60 genes known to be involved in non-syndromic hearing loss (NSHL), and 34 genes known to be associated with 10 forms of syndromic hearing loss (SHL). Nevertheless, it is estimated that there are many more genes to be found implicated in human deafness. Furthermore, there are countless small families that remain unsolved with respect to the cause of their sensory loss. The difficulties of obtaining large pedigrees, the laborious process of genotyping, linkage analysis, positional cloning, and sequencing of dozens of genes in a critical region has hindered the process of further gene discovery. Clinically, this has limited carrier testing of deafness-causing mutations only to the most common ones, or to a specific mutation having been found previously found in the same ethnic population. Therefore the reality is that despite there being over 60 known genes, only one to four genes can be examined for each patient. Since the completion of the Human Genome Project (HGP) and the appearance of Next Generation Sequencing (NGS), or Massively Parallel Sequencing (MPS) and Deep Sequencing, technologies, rapid high-throughput identification of new deafness genes and mutations has been greatly streamlined. A single deep sequencing experiment can be completed within a few days and lead to the identification of new mutations. While the computational analysis and validation of variants still remains the bottleneck for finding new mutations, as more

* Tel: 03-640-6642, Fax: 03-640-9360
genes are discovered and bioinformatics tools developed, this process will become much easier. The application of genomic analysis for early clinical diagnosis of deafness will enable prediction of abnormal phenotypes that might arise later, improve rehabilitation of deaf children and simplify decisions for subsequent treatment. Identification of more genes involved in the auditory and vestibular pathways will help elucidate pathways and mechanisms of inner ear function. In this chapter we will review the advances deep sequencing is having on the discovery of new deafness genes and pathogenic mutations in these genes.

INTRODUCTION

Defining genetic mutations in human genes is unequivocally the first step to determine the pathways involved in a disease, so that targeted therapy may be developed. The general approach to human disease gene identification has been to take a positional cloning approach in order to identify an approximate chromosomal location, and follow with a candidate cloning approach in order to screen relevant genes in the region that might be suitable candidates based on the known phenotype of the disease. Until recently, identification of deafness loci was performed mainly by one of two ways: candidate gene analysis and genome-wide linkage analysis in family members with shared genetic relationships using genetic markers such as microsatellites or single nucleotide polymorphisms (SNPs). As many deafness-associated mutations have been previously identified, an initial strategy is to examine these genes and if they are excluded, to work towards discovery of a new gene. When no known gene is deemed as a potential candidate or when the candidate gene analysis strategy has failed, genome-wide segregation analysis is performed in order to determine the chromosomal region that segregates with a trait. The genetic linkage data obtained could be analyzed by various methods, such as parametric multipoint linkage analysis and, in the case of autosomal recessive disorders, homozygosity mapping, particularly in consanguineous families [1, 2]. As the next step, the causative gene within the genetic linkage data is obtained by various methods. When the size of the linked region is relatively small the conventional strategy has been to utilize Sanger sequencing to sequence the coding regions of each gene in the linked interval one by one in an order based on candidate genes prioritization. For large intervals this technique is costly and time-consuming, making it more necessary to focus on genes known to be expressed in the inner ear. Such methods can be impractical, especially in the case of large intervals as the risk of missing the causative mutation is high.

Over the years the linkage analysis approach has led to the identification of many deafness genes (Hereditary Hearing Loss Homepage, http://hereditaryhearingloss.org/). Deafness may be isolated, as in non-syndromic hearing loss (NSHL), or in conjunction with other abnormalities, as in syndromic hearing loss (SHL). The first NSHL locus was mapped to chromosome 5q31 in 1992 by using linkage analysis in a large Costa Rican family [3]. Mutations in the gene leading to hearing loss (HL), diaphanous 1, were found five years later [4]. The first NSHL recessive locus was mapped in 1994 [5] and three years later, mutations were found in GJB2 encoding connexin 26 [6]. This gene has turned out to be the most common form of genetic deafness, responsible for 30-50% of congenital HL [7]. Since then, the progress in the identification of new genes has been remarkable, with new genes reported at a monthly rate. However, the list of unresolved human loci linked with HL remains longer than the list of cloned genes (Hereditary Hearing Loss Homepage). This is due to the
limitation of linkage methods and the length and cost required for gene identification. Limitations in using linkage methods include the need for relatively large families with multiply affected individuals and a two-step process of determining the chromosomal region of the disease locus and then searching through genes in the region for the causative mutation. Several large-scale screening technologies have emerged that increase facilitated disease gene identification for deafness. One such example is the use of array comparative genomic hybridization (array CGH) for the identification of chromosomal imbalances such as genomic deletions and duplications. Using this method, an inverted genomic duplication of the TIP2 gene was discovered to be the cause of progressive NSHL in DFNA51 [8]. A second example is the Hereditary Hearing Loss Arrayed Primer Extension microarray that enables an analysis of 198 mutations across eight prevalent genes in a single test (GJB2, GJB6, GJB3, GJA1, SLC26A4, SLC26A5, MTRNR1 and MTTS1) [9]. Such a tool can be useful in the clinic; however it does not detect new mutations or mutations within genes not previously associated with deafness. Though the advances in deafness gene discovery have been considerable, the great limitations that are still encountered indicate a need for large-scale techniques that will screen a larger number of genes in a reasonable amount of time and more cost-effective manner, and that can detect all types of mutations underlying deafness.

**THE EMERGENCE OF NGS TECHNOLOGY**

Sequencing of the first human genome was performed using the chain-terminator method, also known as Sanger sequencing [10]. The first diploid sequence of an individual was reported in 2008, that of J. Craig Venter, the founder of Celera Genomics, with 1.5 gigabase (Gb) of sequence, performed at an estimated cost of 200 million dollars [11]. However, Sanger sequencing has limitations, as it is time- and cost-consuming to perform. The number of large genes involved in HL has made complete sequencing on a regular basis impractical, even for research purposes, and the costs of identifying new deafness genes are tremendous. Next Generation Sequencing (NGS), also named Massively Parallel Sequencing (MPS) or Deep Sequencing, enables large scale screening of a wide variety of genes and in addition, identification of mutations in non-coding regions that may affect gene expression. While Sanger sequencing yields an output of 120,000 bp per day for a cost of approximately $4000 per Mb sequenced [12], with a cost of $200,000 to sequence 3.2 Gb of a single human genome, the output of a single MPS machine stood at 30 Gb per day in 2011 [13], with the costs coming down almost daily. The genome of James Watson, winner of the 1962 Nobel Prize in Physiology or Medicine for his co-discovery of the molecular structure of nucleic acids (http://www.nobelprize.org/nobel_prizes/medicine/laureates/1962/), was sequenced using this newer technology and ran at a cost of approximately $1 million, a far cry from the cost by Sanger sequencing [14]. When compared to Sanger sequencing, MPS has an increased error rate, due to the increased output [15], but this problem is overcome by repeated sequencing to achieve greater coverage or read depth. Currently there are three commonly used commercial MPS platforms. These include the 454 Sequencing System (454 Life Sciences, a Roche company), the first MPS platform introduced, using pyrosequencing [16]; the Illumina platform (Illumina, Inc.), the most widely used platform, using cyclic reversible termination (CRT) technology [17]; and the SOLiD™ System (Life Technologies, Applied Biosystems) using sequence-by-igation (SBL) technology [18]. Each of the
technologies mentioned has advantages and disadvantages (for a comprehensive review, see [13, 19]). One of the major challenges of the NGS technology is the formidable amount of data generated, thus making it necessary to develop bioinformatic filtering techniques, data analysis tools and storage solutions. Adding a genomic enrichment step to isolate a selected region of the genome can greatly reduce the costs and bioinformatic infrastructure necessary for data analysis. Narrowing the studied genomic region can be done by either targeted genomic capture or by whole exon (exome) sequencing. In targeted genomic capture, any region of interest in the genome can be targeted including disease-associated linkage intervals or the exons of genes that are known to be associated with a specific disease or organ. There are two main methods for targeted capture, solid-phase targeted enrichment, based on a DNA microarray technology, and liquid-based targeted enrichment, using RNA or DNA based complementary probes (for a comprehensive review of targeted capture, see [20]). Whole exon sequencing enriches all the exons of the genes within the genome for NGS. One of its major advantages is that it does not limit the genes analyzed to ones already associated with the disease, thus allowing for the identification of new disease genes.

NGS technologies have revolutionized the field of gene discovery and genetic testing by enabling the detection of genomic variations within millions and even billions of base pairs. In many cases these variations are benign polymorphisms with no clinical significance. One of the challenges of analyzing data generated by NGS technology is to define the significance of the variants detected and decide which are SNPs and which are disease-causing mutations. Approaches to rule out SNPs include searching large sequencing projects (such as the 1000 Genomes Project, http://1000genomes.org), the NHLBI Exome Sequencing Project (http://evs.gs.washington.edu/EVS/), and databases of known polymorphisms (such as the dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/), applying computational prediction of the potential effect of the variation on protein function, and sequencing the gene containing the variation in a large set of control population from a similar genetic background. Such a step is extremely important to detect the variants most likely to be disease-causing mutations.

NGS in Hearing Loss Research

NGS offers an ideal solution for the many challenges in the identification of causative mutations in heterogenic diseases such as deafness. Since its emergence, NGS has had an important impact on both research and clinical diagnosis in many diseases including cancer [21-24], inflammatory bowel disease [25], and cognitive disorders [26], as it enables screening of a large number of genes in one test. The many advantages of this system to deafness research include parallel analysis all known deafness genes; identification of mutations in genes previously not implicated in deafness; detection of all types of mutations; screening of large genes that have previously been largely untested; parallel processing of a large number of samples; and the fact that it can be used in cases of isolated deafness. It is therefore not surprising that in the last two years this technology has begun to play a pivotal role in elucidating the molecular genetic basis underlying deafness. To date, the two methods used in this field have been targeted gene capture and multiple parallel sequencing or homozygosity mapping prior to exome sequencing (for an example of an experiment scheme see Figure 1).
Figure 1. A representative scheme of a NGS experiment. This scheme illustrates the workflow utilized to identify GPSM2 as the cause of deafness. (A) The investigated Family CG with inheritance of a mutation in GPSM2 and hearing loss (filled symbols). (B) A diagram illustrating the workflow used to identify GPSM2 as the causative mutation in Family CG. A combined homozygosity mapping and whole exome sequencing approach was utilized in order to narrow down the data that needed to be analyzed. In the region analyzed, 80 variants were detected, 73 of which appeared to be known SNPs. Of the seven variants previously unidentified, five were sense mutations and two were amino acid (AA)-altering mutations, within the MYBPHL and GPSM2 genes. The two variants were analyzed in a large control population. While the variant in the MYBPHL gene was identified in one percent of the control population, the substitution in the GPSM2 gene was not present in the control population. Further analysis found that the mutation in GPSM2 segregates with deafness in Family CG. (C) Immunohistochemistry staining for Gpsm2 within the e16.5 and P0 mouse inner ear. Gpsm2 is expressed at the apical surface of the hair and supporting cells of the cochlea, utricle, saccule and cristae. Expression in the cochlea is shown. Adapted from [46].
This more targeted approach is due to the current bioinformatics limitations, as well as time and budget constraints. However, in the near future such restrictions will surely decrease, thus allowing a more widespread use of whole-exome or whole-genome sequencing. Within the field of human deafness research, to date, NGS has been predominantly used for two major applications: the discovery of new deafness causative genes and new mutations within known genes; and large-scale screening of affected individuals and generation of clinical diagnosis platforms. These two approaches enable both the expansion of the understanding of deafness genetics, as well as clinical application of the current knowledge of genes underlying deafness.

**NGS FOR DISCOVERY OF NEW DEAFNESS-CAUSATIVE GENES AND PATHOGENIC MUTATIONS**

NGS can be a powerful tool in the detection of new deafness-causing mutations, both in previously known genes, as well as in genes not previously implied in HL. This technology helps overcome the serious drawbacks of linkage analysis, the most widely used method to search for genetic disease mutations in the genomic era. Indeed, since this technology entered the market, its use to identify causative genes and mutations in deaf individuals has grown tremendously. Targeted capture and MPS appears to be the ideal tool for this task as it limits the analysis to a manageable subset of the whole genome, thus reducing the cost, time and bioinformatics infrastructure necessary for the analysis of large-scale data.

**COMBINED TARGETED CAPTURE AND MPS APPROACH**

The first study utilizing targeted genomic capture and MPS within the field of HL research identified the C9orf75 (TPRN) gene, encoding taperin, as the mutated gene in NSHL DFNB79 [27]. Targeted capture was conducted on a refined region containing 108 genes, with 870 protein-coding exons of the previously identified DFNB79 linkage region [28]. Protein-coding sequences, as well as untranslated regions (UTRs), introns and noncoding genes were analyzed. The genomic DNA (gDNA) of a single affected individual from a Pakistani consanguineous DFNB79-linked family was analyzed, revealing a nonsense mutation in the TPRN gene. Subsequently, mutations within the additional three DFNB79-linked families were identified as frameshift mutations within the TPRN gene. Furthermore, using immunolocalization, within the mouse cochlea, the mouse Taperin orthologue was localized at the taper region at the base of the hair cell stereocilia, thus providing the gene’s descriptive name. In addition, taperin was localized to the supporting cells and to a lesser extent to Reissner’s membrane and spiral ligament. These findings led to the speculation that taperin may modulate actin dynamics through direct or indirect interaction with pointed ends of actin filaments.

Two studies followed that utilized targeted genomic capture and MPS technology to identify the X-linked gene SMPY, encoding small muscle protein, as the causative gene for DFNX4 nonsyndromic hearing impairment. Genome-wide linkage analysis was utilized on 11 affected individuals from a large German family with X-linked NSHL to reveal linkage to a
17.5 MB region on chromosome Xp22.12 overlapping the DFNX4 locus previously identified in a study of a Spanish family [29, 30]. Targeted genomic capture with enrichment of all exons and 1 kb of the promoter regions of the 88 coding genes within the linked interval was conducted followed by NGS. Six variants located in exons that were not annotated as SNPs were identified. The nonsense mutation in the SMXP gene was considered the best candidate and was validated using Sanger sequencing. Subsequent investigation of the DFNX4-linked Spanish family led to the discovery of a nonsense mutation within the SMXP gene that segregates with HL in the family. In parallel, linkage analysis for the known X-chromosomal NSHL loci, DFNX1-5 led to the identification of a critical linkage interval overlapping the DFNX4 locus [29, 31]. Using NGS and filtering for predicted sequence variants against dbSNP, the 1000 Genome Project and 200 Danish controls, a single variant within the linkage interval, a nonsense mutation in the SMXP gene that co-segregated with hearing impairment in the family, was detected. Subsequent Sanger sequencing-based screening of 26 individuals from small families for which X-linked NSHL was not excluded revealed a frameshift mutation in the SMXP gene in one of the screened individuals. SMXP was initially cloned from human muscle [32], where it localizes to costamere structures that protect the sarcotendinous plasma membrane from damage generated by the mechanical stress caused by the contraction of muscle cells [33]. Using both quantitative PCR on fetal human inner ear [31] and immunohistochemistry of the adult mouse inner ear [30], the expression of SMXP was verified within the mammalian inner ear. Within the organ of Corti, SMXP was localized to different cell types, including Böttcher cells, root cells, pillar cells, and interdental cells of the spiral limbus, with low levels of expression in the hair cells [30].

In addition to its use in the discovery of deafness genes causing NSHL, a combined approach using targeted capture and MPS has been used to identify causative mutations leading to SHL. To date, mutations leading to SHL were identified in 36 genes leading to 10 syndromes (Hereditary Hearing Loss Homepage). These mutations are both in nuclear and mitochondrial DNA. Though quite a few mutations have been identified thus far, there are still many families with SHL for whom the causative gene mutated gene is still unknown. In an attempt to uncover the genetic basis of Perrault syndrome, in a non-consanguineous European family, a combined linkage analysis and Sanger sequencing approach together with NGS was undertaken [34]. Perrault syndrome is a genetically heterogeneous recessive disorder characterized by sensorineural hearing loss and ovarian dysgenesis, as well as neurological manifestation in some families [35-37]. Genome wide linkage analysis identified a 4.142 Mb interval on chromosome 5 containing 58 genes. The exons of these genes, as well as the flanking regulatory regions, were analyzed using Sanger sequencing and missense mutations were found in the HARS2 gene in affected individuals. In order to rule out additional candidate mutations within the linked region, target capture and MPS were conducted and HARS2 was identified as the only gene in the regions with variants leading to a predicted functional affect. In addition to the missense mutation, in one of the two alleles the nucleotide substitution created an alternative splice site, leading to a 12 codon deletion. HARS2 encodes a mitochondrial histidyl tRNA synthetase [38] that catalyzes the covalent linkage between tRNAs and specific amino acids, a process necessary for the translation of mitochondrial-encoded proteins [39]. Analyzing the mutations in yeast and Caenorhabditis elegans, Pierce et al. were able to conclude that Perrault syndrome was caused by a reduction in HARS2 activity in the family studied.
Mutations in the \textit{DNMT1} gene were identified in two separate studies using NGS technology for two distinct syndromes that include HL in their phenotype. \textit{DNMT1} encodes the DNA methyltransferase 1 protein, a widely expressed protein that maintains methylation patterns in development, and mediates transcriptional repression [40]. Utilizing linkage analysis and exome sequencing, a form of hereditary sensory and autonomic neuropathy type 1 (HSAN1) syndrome, a neurodegenerative disorder with central and peripheral involvement including dementia and hearing loss [41], was studied. The form investigated in the study showed early onset dementia and sensorineural hearing loss, followed by sensory neuropathy. Four kindreds with similar phenotypes were investigated, two American, one Japanese and one European. In all cases one of two nonsense mutations were identified in exon 20 of the \textit{DNMT1} gene within the target-sequencing domain that regulates DNMT1 binding to chromatin, leading to premature degradation of the mutant protein. Using \textit{in vitro} studies, the mutations were demonstrated to disrupt proper folding of the protein, preventing it from entering the nuclei, as well as an effect on the stability and enzymatic activity of the full length protein, leading to global hypomethylation and local sight-specific hypermethylation. Another study performed whole exome sequencing in five affected individuals from three different kindreds with autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN) characterized by late onset sensory neuronal deafness, narcolepsy or cataplexy, cerebellar ataxia and dementia [42]. The studied population included an American family, a Swedish family and Italian kindred. After filtering for SNPs, only a single gene, \textit{DNMT1}, was mutated in all five individuals studied, with two different mutations detected. In all cases the mutations were in exon 21. Sanger sequencing confirmed the mutations and showed cosegregation for the American and Swedish mutations with affected individuals. Sanger sequencing was also used to identify an additional mutation within exon 21 of the \textit{DNMT1} gene in an ADCA-DN Italian family.

\section*{Whole Exome Sequencing in HL Research}

Whole exome sequencing screens the exons of all genes in the human genome, thus enabling the discovery of novel deafness genes. It is estimated that approximately 60\% of genes for Mendelian disease could be discovered using this technology [43]. This method can be extremely useful in cases involving small or non-consanguineous pedigrees where neither linkage analysis nor homozygosity mapping would be informative. In one study, whole exome sequencing of genomic DNA from one of two sisters with Perrault syndrome identified two rare functional variants in the \textit{HSD17B4} gene, leading to a compound heterozygote genotype [44]. \textit{HSD17B4} encodes 17b-hydroxysteroid dehydrogenase type 4, a peroxisomal enzyme involved in steroid metabolism and fatty acid $\beta$-oxidation. In this study showed that one of the two variants uncovered, a missense mutation, is predicted to destabilize the HSD17B4 dehydrogenase domain, whereas the second, a nonsense mutation, leads to very low levels of the HSD17B4 transcript. Furthermore, they were able to show that expression of mutant HSD17B4 protein in a compound heterozygote was severely reduced. Thus using NGS, this study was able to discover mutations in two separate genes, \textit{HSD17B4} and \textit{HARS2}, as the cause of Perrault syndrome in two families [34, 44]. These are the only two genes thus far linked to Perrault syndrome and without the NGS technology these
mutations would most probably have not been identified. The authors analyzed HSD17B4 and HARS2 were analyzed in six additional families and another study analyzed these genes in eight Perrault syndrome families [45]. However, no additional mutations were detected, indicating that other critical genes remain to be identified.

The major limitation of whole exome sequencing is in the vast bioinformatics infrastructure necessary for data analysis. One approach to reduce this limitation is to compare variants within the exome after filtering for known SNPs, as utilized to identify mutations in DNMT1 [42]. Another approach is to perform a whole exome sequencing experiment along with homozygosity mapping for recessive families in order to narrow down the regions to analyze. Three studies utilized such a method, the first of which was a study that identified a mutation in the gene GPSM2 as the cause of NSHL DFN828 in a consanguineous Palestinian family (Figure 1A-B) [46]. The DFN82 region, located on chromosome 1, spanned 50 annotated protein-coding genes and one microRNA. Whole exome sequencing was conducted on genomic DNA from a single affected individual. By focusing on exonic and flanking intronic variants within the linked region, the data generated was narrowed down (Figure 1). To further define the potential role of Gpsm2 in the inner ear, immunohistochemistry staining for Gpsm2 was performed, revealing its expression in a pattern characteristic for proteins involved in the planar cell polarity pathway. A second truncating mutation was found in the GPSM2 gene in a consanguineous Turkish family [47]. A subsequent study revealed that these patients previously diagnosed with NSHL also had Chudley-McCullough syndrome (CMS), with frontal polymicrogyria, abnormal corpus callosum and gray matter heterotopia [48]. Homozygosity mapping was also used to narrow down the data generated by whole exome sequencing for three individuals from two consanguineous Turkish families with phenotypes resembling the Carnevale, Malpuech, Michel, and ocu-sole-skela-abdominal (OSA) syndromes, including distinctive facial features, skeletal anomalies, developmental delay, periumbilical defects and mixed hearing loss [48]. Whole exome sequencing of one proband and analysis of the mapped region revealed a missense mutation in the MASP1 gene, encoding mannann-binding lectin serine protease 1, which co-segregated with the phenotype. In the second family a nonsense mutation in the MASP1 gene was identified using Sanger sequencing. As the two mutations occur in a MASP1 isoform previously reported to process insulin-like growth factor binding protein (IGFBP5) [49], the authors concluded that MASP1 most probably plays a critical role in insulin growth factor availability during craniofacial and muscle development.

A third study combining homozygosity mapping and whole exome sequencing identified two novel missense variants in GIPC3, encoding a protein that localized to the inner ear hair cells, and ZNF57, a zinc finger gene, [50]. Both variant co-segregated in affected individuals with NSHL born to a consanguineous Turkish family. An initial analysis of the exome sequencing data using a read depth filter of ≥8x identified only the mutation in ZNF57. However, while the study was ongoing, reports of two missense mutations in GIPC3 in two small families with sensorineural HL and an association between a mutation in GIPC3 and age-related sensorineural HL in mice were published. The authors re-analyzed their data using a less stringent read depth filter of ≥4x. This analysis detected two additional variants within the autozygous region with only one, in exon 3 of the GIPC3 gene, confirmed by Sanger sequencing. The authors concluded that the HL in the family they describe is caused by a novel missense mutation in GIPC3, as such mutations have been described in other families with NSHL. They speculated that it is possible that the mutation in ZNF57 may
contribute to the phenotype or to susceptibility to another phenotype that was not considered in their study. This work is a clear example of a major challenge of using NGS technology for the discovery of causative mutations in Mendelian diseases. It shows that even when a whole exome study attains the expected quality standards for the number of reads, the amount of DNA generated, the target coverage and the average depth across the intended target, a marked unevenness of capture of one region may occur, influencing the results.

One common approach used to identify disease-causing mutations within a genetically linked region is to search for variations in genes within this region that were previously shown to either cause deafness or to be expressed within the inner ear. In such a case one must exclude additional potential disease-causing variants within the linked region. In the past, this approach has been done using Sanger sequencing. However, such a method is extremely time- and cost-consuming, especially in the case of a relatively large linked region. NGS was used to verify that a mutation in CEACAM16 is the cause of autosomal dominant NSHL in an American family linked to the DFNA4 locus [51]. CEACAM16 is a member of the carcinoembryonic antigen-related cell adhesion family of adhesion molecules [52]. CEACAM16 was shown to be expressed in the cochlear outer hair cells, the tips of the tallest stereocilia and the tectorial membrane, and to interact with α-protectorin, a major component of the tectorial membrane. Thus it was a candidate for a deafness gene based on chromosomal location and predicted function. Indeed, mutation screening of its coding exons revealed a heterozygous missense mutation. This mutation co-segregated with the phenotype within the family and was not identified in ethnically-matched controls. The DFNA4 interval contains 715 genes, therefore, in an attempt to exclude variants in other genes within this region, whole exome sequencing was performed on genomic DNA from a single affected individual. After filtering for documented SNPs, the authors identified 12 additional variants within the linked interval. However, as these were either not verified by Sanger sequencing or did not co-segregate with HL, they were excluded, providing strong evidence that the mutation in CECA16 is the cause of HL within this family.

NGS FOR SCREENING AND CLINICAL DIAGNOSIS PLATFORMS

To date, PCR, restriction enzyme analysis and Sanger sequencing are the most widely used technologies for the detection of genetic mutations in the clinics. Routine clinical use of NGS technologies in the clinical setting is appealing as it has the potential to decrease sequencing costs and increase throughput. However, translational use in clinical care mandates high accuracy, simple assays, inexpensive instruments, short run times and easy data analysis. In a few cases, NGS has been used clinically for diagnostic purposes, especially in cancer [53, 54]. Hereditary HL can be predicted by genetic testing with high accuracy as the cause is monogenic in most cases, thus making it an ideal candidate for genetic screening using NGS. Indeed, in the last two years, a number of studies have set out to create platforms for diagnostic mutation detection of focused panels of deafness genes.

The feasibility of target enrichment and MPS technologies was used to examine the exons of 54 genes associated with human NSHL [55]. The genes analyzed included Usher syndrome genes since the Usher syndrome phenotype is not always distinguished from NSHL in children. The platform and diagnostic pipeline was named OtoSCOPE (Otopologic Sequence Capture Of Pathogenic Exons). The study compared two platforms, the SureSelect-Illumina
(solution-based) and NimbleGen-454 (solid-phase) platforms, in an effort to determine the most efficient method for identifying deafness genes for screening towards clinical diagnosis. They tested nine individuals diagnosed with genetic HL, including three positive controls diagnosed by Sanger sequencing prior to the study. Pathogenetic mutations in the STRC, CDH23, MYO6, KCNQ4 and MYH14 genes were found in six probands. While both platforms provided high specificity and sensitivity, the authors concluded that under their conditions the SureSelect-Illumina platform was preferable with regards to scalability, sensitivity, and cost. In a further effort to develop clinical diagnosis platforms, a combined targeted genomic capture and MPS on a cohort of Israeli Jewish and Palestinian Arab hearing impaired probands was performed [56]. In this study, a custom cRNA oligonucleotide design containing 246 genes responsible for syndromic and NSHL in humans and mice was constructed. Targets were 82 human-protein coding genes, 2 human microRNAs and the human orthologues of 162 mouse deafness genes, based on the fact that many of the human deafness genes also cause deafness in mice. Using this array, samples from 11 probands of families with HL were sequenced. After filtering for reported SNPs, rare single base pair and insertion/deletion (indel) variants were identified. Using this approach, critical mutations were identified in six of the original probands and their families. These include mutations in the TMC1, CDH23, MYO15A, TECTA and WFS1 genes. These mutations were found to co-segregate with HL in the affected families. A novel mutation identified in TMC1 was further shown to contribute to 38% of hereditary HL in the Moroccan Jewish population, solving the cause of HL in an additional 20 families. This approach exploits the high-throughput nature of targeted MPS to make a single fully comprehensive test for all known deafness genes. The approach had the added benefit of including genes associated with deafness in mice, thereby having the potential to discover additional human deafness genes.

One of the main limitations of an NGS-based platform for clinical application is the high per-sample cost. In an attempt to reduce costs, in-house cDNA-based probes were produced for targeted gene enrichment [57]. As proof of principle, PCR amplicons from cDNA clones of five genes known to cause deafness (GJB2, GJB3, GJB6, SLC26A4 and MYO15A) were used as bait probes in hybridization for capturing human genomic DNA (gDNA) fragments. The captured gDNA fragments were then sequenced using the Illumina GAII platform. The authors used gDNA samples from two normal hearing human subjects and ten individuals previously shown to have genetic mutations in GJB2 by Sanger sequencing. The results were confirmed by Sanger sequencing in all cases. The authors concluded that their approach achieves the necessary specificity, depth of coverage, multiplexicity and uniformity for accurate sequencing by a NGS platform, as well as significantly reduces per-patient cost.

**CONCLUSION**

NGS has transformed genomic research by decreasing the costs of sequencing and increasing throughput. Its impact on the auditory field is in its infancy and will no doubt increase by many fold in the near future. Applications of NGS technology have recently begun to transition into clinical care. Such applications will surely enable rapid pathogenic mutation screening that will be widely used, thus reducing the need for multiple parallel diagnostic approaches. In addition, there is great hope that NGS will greatly advance patient
tailored therapeutics. The advances made so far are encouraging, but for wide and routine use in the clinic, additional improvements will need to be made, including reduction in costs, simple assays and easy data analysis. A number of NGS systems recently launched have some of these characteristics, but still face a number of challenges [58]. In addition, ethical issues will need to be addressed; including the possibility that this new technology may exacerbate the ambivalence society has towards physical differences and anomalies [59]. Furthermore, the likelihood of unintended discoveries as a by-product of the technology complicates matters even further. Therefore the issue of an individual's genomic privacy is of utmost importance and invites questions such as, who has accessibility to the data generated and how is it stored and disposed of, which must be promptly addressed [60]. Finally, the impact deep sequencing has had on the hearing and deafness field is dramatic. The ethical implications described above play a role in this field as well. Furthermore, there is the added issue that termination of pregnancy was controversial in this area [61], and once genetic diagnosis is available to all, the option of using it for prenatal diagnosis and family planning is a reality. In addition, deep sequencing may identify more than one variant that appears to be a mutation without a sufficient number of affected family members to provide validation, making it difficult to conclusively determine which is the causative mutation and complicating genetic counseling. Nonetheless, the ability to diagnose and define a mutation has become much more amenable for the hearing impaired population. This has and continues to pave the way for elucidation of pathways and mechanisms of genetic hearing loss, with the impending development of therapeutics.

ACKNOWLEDGMENTS

Research in the Avraham laboratory for human genomics is funded by the National Institutes of Health (NIDCD) R01DC011835; Israeli Centers of Research Excellence (I-CORE), Gene Regulation in Complex Human Disease, Center No 41/11; and the Hedrich Charitable Trust. We thank Shaked Shivatzki for help with figures.

REFERENCES


