Published in final edited form as:

Hear Res. 2012 June ; 288(0): . doi:10.1016/j.heares.2012.01.004.

Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities

Xi Lin^{1,2}, Wenxue Tang¹, Shoeb Ahmad¹, Jingqiao Lu¹, Candice C. Colby¹, Jason Zhu¹, and Qing Yu^{1,3}

¹Department of Otolaryngology, Emory University School of Medicine, 615 Michael Street, Atlanta, GA 30322-3030

²Department of Cell Biology, Emory University School of Medicine, 615 Michael Street, Atlanta, GA 30322-3030

³Beijing Tongren Hospital, #2 Chong-wen-mennei Street, Dongcheng District, Beijing, 100730

Abstract

The goal of sequencing the entire human genome for \$1,000 is almost in sight. However, the total costs including DNA sequencing, data management, and analysis to yield a clear data interpretation are unlikely to be lowered significantly any time soon to make studies on a population scale and daily clinical uses feasible. Alternatively, the targeted enrichment of specific groups of disease and biological pathway-focused genes and the capture of up to an entire human exome (~1% of the genome) allowing an unbiased investigation of the complete protein-coding regions in the genome are now routine. Targeted gene capture followed by sequencing with massively parallel next-generation sequencing (NGS) has the advantages of 1) significant cost saving, 2) higher sequencing accuracy because of deeper achievable coverage, 3) a significantly shorter turnaround time, and 4) a more feasible data set for a bioinformatic analysis outcome that is functionally interpretable. Gene capture combined with NGS has allowed a much greater number of samples to be examined than is currently practical with whole-genome sequencing. Such an approach promises to bring a paradigm shift to biomedical research of Mendelian disorders and their clinical diagnoses, ultimately enabling personalized medicine based on one's genetic profile. In this review, we describe major methodologies currently used for gene capture and detection of genetic variations by NGS. We will highlight applications of this technology in studies of genetic disorders and discuss issues pertaining to applications of this powerful technology in genetic screening and the discovery of genes implicated in syndromic and non-syndromic hearing loss.

© 2012 Elsevier B.V. All rights reserved.

Correspondence: Xi Lin, PhD Department of Otolaryngology and Cell Biology Emory University School of Medicine Atlanta, GA 30322 Telephone: 404-727-3723, Fax: 404-727-6256 xlin2@emory.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure Statement Drs. Xi Lin and Wenxue Tang own equity in the Ototogenetics Corporation, which offers a commercial service for capturing and next-generation sequencing of human genes implicated in syndromic and non-syndromic hearing loss. The terms of the conflict of interest (COI) arrangement for the two authors have been reviewed and approved by Emory University in accordance with its COI policies.

Keywords

mutations; genes; deafness; genetics; review; next-generation sequencing; exome; gene enrichment

Introduction

Innovative application of new technologies in research is one of the major factors driving advances in knowledge acquisition. Landmark events in DNA studies include the Sanger sequencing method introduced in 1977 (Sanger et al., 1977) and discovery of the polymerase chain reaction (PCR) in 1983 (Inoue et al., 1983), which enabled many applications in molecular analysis of DNA and genes. A common pattern in their application is that the commercialization of these technologies first led to widespread uses in the research community, and later into the clinical diagnosis of many genetic disorders (e.g., <http://www.ncbi.nlm.nih.gov/sites/GeneTests/>). Based on the Sanger technology, the first human genome was sequenced in 2001 by a multinational collaborative effort of the Human Genome Project, which took 13 years to complete at a cost of nearly USD \$3 billion (Lander et al., 2001; McPherson et al., 2001; Sachidanandam et al., 2001). With the commercial availability of high-throughput massively parallel DNA sequencing platforms in the past few years, complete sequencing of the whole human genome can be done commercially today in two to three months at a cost below \$10,000 (Bick et al., 2011).

A major focus of this review is a description of the combined use of gene capture followed by next-generation sequencing (NGS) in applications for screening mutations in known disease genes. Since their introduction, NGS technologies have constantly improved and the costs have steadily decreased. A legitimate question therefore is what role targeted gene capture will play if whole-genome sequencing (WGS) can be done for about \$1,000 in the future, with the ability to survey the human genome in an unbiased manner. A simple answer for the continued use of targeted gene capture prior to NGS is that the popular description of “\$1,000/human genome sequencing” is only a fraction of the total cost of a meaningful application of such a technology. WGS requires the analysis of $\sim 3.2 \times 10^9$ base pairs (bps) of DNA sequences. The whole human exome (all protein-coding regions of the human genome, which is approximately 1% of the genome) and a targeted panel of genes involved in deafness cover 30–60 million bps (Clark et al., 2011) and ~ 0.3 million bps (supplemental Table 1), respectively. Adding a targeted capture step prior to NGS enables the analysis of a focused subset of the whole genome, thereby reducing the required sequencing capacity and corresponding bioinformatic infrastructure necessary for data management and analysis. Vast amounts of genotype and phenotype data about human diseases are now being generated. The bioinformatic filtering techniques, storage facilities, software, and hardware for data analysis, especially the data annotation with unequivocal functional interpretations, are formidable challenges for the meaningful application of WGS.

Given a set amount of NGS capacity and budget, a significantly larger number of samples can be processed with the targeted NGS approach. This results in a greater ability to detect variations on a population level, which is a key requirement for discovering new causative variants and for verifying benign single- and multiple-nucleotide polymorphisms (SNPs and MNPs) in human populations. In addition, most clinical applications of NGS concentrate on known mutations in order to generate clear interpretable reports. The combined uses of capturing a targeted panel of disease genes followed by NGS apparently suit this aspect of clinical study. Targeted sequencing has been shown to be a robust, effective technique that leverages the unique aspects of massively parallel sequencing and has already yielded many exciting discoveries (Brownstein et al., 2011; Majewski et al., 2011; Shearer et al., 2010).

This review focuses on what we can learn from the combined application of gene capture and NGS regarding human diseases, especially in studies of human genes implicated in human syndromic and non-syndromic hearing loss. The current state of gene capture and NGS technologies will be presented. Many practical aspects of performing targeted gene capture, WES (whole-exome sequencing), and WGS in particular applications will be discussed.

Overview of massively parallel NGS technologies

The Human Genome Project was completed largely based on the Sanger method, which demonstrated that the method is costly and time-consuming for large-scale studies of many genes in parallel. NGS has dramatically reduced the cost of sequencing on a per-base pair (bp) basis and increased the output of sequencing from a few hundred bps by each Sanger analysis to about 600 billion bps per NGS machine run (Clark et al., 2011; Zhang et al., 2011).

NGS is carried out by sequencing small DNA fragments in a parallel fashion. The first step in NGS is carried out by fragmenting the genomic DNA (mechanically or chemically) into small pieces, usually in the range of 300-500 bps (Borgstrom et al., 2011). In the next step, platform-specific adapters are added to the ends of the DNA segments, allowing for their attachment and sequencing. One common feature shared by almost all current NGS platforms is that clonally amplified single DNA molecules, spatially separated in a defined microchamber (called Flow Cells, FlowChips or PicoTiter plate), are sequenced in a massively parallel fashion. One exception is Pacific Biosciences' single-molecule method, which has a protocol based on real-time sequencing without clonal amplification (Eid et al., 2009).

In the NGS execution, sequencing results are generated by reading optical signals during repeated cycles from either polymerase-mediated fluorescent nucleotide extensions of four different colors (Ruparel et al., 2005) (e.g., Illumina's HiSeq system), or from iterative cycles of fluorescently-labeled oligonucleotide ligation (e.g., ABI SOLiD system), or by the principle of pyrosequencing (Margulies et al., 2005) (e.g., Roch 454 system). Non-optical DNA sequencing by directly sensing the hydrogen ions produced by template-directed DNA polymerase synthesis on semiconductor-sensing ion chips has recently been developed as well (Rothberg et al., 2011). In such a massively parallel sequencing process, NGS platforms are able to generate up to 600 gigabases of nucleotide sequence from a single instrument run (e.g., Illumina's HiSeq200 V3 chemistry kit) (Clark et al., 2011; Zhang et al., 2011). The results of the sequenced segments are called "reads," which could be 25-100 bps from one or both ends. The massive capacity of NGS allows the sequencing of many randomly overlapping DNA fragments; therefore, each nucleotide in targeted regions may be included in many reads, allowing repeated analysis which provides depth of coverage. Increased depth of coverage usually improves sequencing accuracy, because a consensus voting algorithm is used in determining the final nucleotide calls. Some recent reviews are available for exploring technical details about the specifics of NGS technologies (Ng et al., 2010a; Shendure et al., 2008).

Overview of targeted gene capture technologies

During the NGS process, hundreds of millions of these DNA segments are sequenced at the same time, in cycles from one or both ends. Without selective enrichment, any genomic region has an equal chance of being sequenced. With targeted gene capture, the proportion of DNA fragments containing or near targeted regions is greatly increased. Any region of interest in the genome can be targeted for enrichment by design, including exons from a list of genes that are known to be associated with diseases (e.g., known hearing-loss genes) or

specific functionally important biological pathways, or disease-associated chromosome segments (including both exons and introns) suggested by linkage analysis. Additional regions of interest, such as promoters or highly conserved sequences in intron regions, can also be targeted for capture.

Currently, at least 156 human genes implicated in syndromic and non-syndromic hearing loss are confirmed (supplemental Table 1), and 57 are non-syndromic hearing loss genes (not including mitochondrial or miRNAs) (Brownstein et al., 2011; Hilgert et al., 2009). Approximately 90 G bp of sequencing is needed to obtain NGS results for one human genome (~3.2 G bp) at about 30x average coverage. The combined mRNA length of the genes listed in the supplemental Table 1 is about 0.01% of the human genome. Targeted gene enrichment typically increases this proportion by at least 1,000 fold (Brownstein et al., 2011; Shearer et al., 2010; Tang et al., 2011). Therefore, the same sequencing capacity can theoretically be used to sequence more than 1,000 samples for a panel of genes associated with deafness listed in the supplemental Table 1. DNA samples collected from the different subjects can be distinguished in NGS reads by adding unique short oligos (6-10 bps) as barcodes when sequencing adaptors are ligated (Craig et al., 2008). Barcoded samples can be pooled and sequenced together, enabling a significant reduction in the per-sample cost of each NGS machine run. Current target enrichment techniques can be generally classified as hybridization-based and non-hybridization methods.

Hybridization-based enrichment methods

Among many current methods for targeted enrichment of specific genome regions, the hybridization-based method is the most widely adopted and is the only one that has been extended to capture the whole human and mouse exome reliably (Clark et al., 2011; Majewski et al., 2011). Such an approach can be classified as liquid-based hybridization (Bainbridge et al., 2010) or solid-phase based (e.g., array-based) hybridization (Choi et al., 2009; Hodges et al., 2007; Ng et al., 2009), depending on how the reactions are experimentally implemented. The former is much more popular now due to the advantages of protocol implementation in automated robotic systems.

(1) Enrichment performed in liquid-phase hybridization—Biotinylated DNA or RNA probes are used in liquid-phase hybridization protocols for target enrichment. Both biotinylated RNA probes (e.g., Agilent's SureSelect exome kit) and DNA probes (e.g., Roche/Nimblegen exome V2 or Illumina TruSeq exome kits) are commercially available. The probes are usually 120 bps long DNA or RNA oligos and have sequences complementary to the targeted genome regions. A tiling scheme which allows overlapped probes to target the region of interest is generally used to increase capture efficiency. Long capture probes made from cDNA clones of the targeted genes are also successfully used (Tang et al., 2011). Hybridization reactions usually take two days to complete. The targeted DNA fragments bound by biotinylated DNA or RNA probes in the solution are captured by magnetic streptavidin beads immobilized by a strong magnet. DNA fragments not targeted are in the liquid phase, and are removed by repeated washes. Targeted DNA fragments are eluted from the beads by increasing the pH with NaOH in order to break the biotin-streptavidin bond and degrade the RNA probes. Only the enriched DNA fragments are sequenced by NGS protocols. The probes used for enrichment do not have the appropriate adaptors on the ends of the DNA fragments and are not sequenced.

(2) Enrichment performed in solid-phase hybridization—Solid-phase hybridization methods generally utilize DNA probes affixed to a solid support, such as microarray glass slides (Albert et al., 2007; Hodges et al., 2007; Okou et al., 2007), filter papers, or films (Herman et al., 2009). Fragmented DNA in a small volume is applied to the surface of the

solid support, where the targeted DNA fragments hybridize to the probes designed to be complementary to the target sequences. The non-targeted fragments will not hybridize and are washed away. The enriched DNA is eluted and used for NGS downstream protocols. Such a method can usually enrich targets by about 1,000-2,000 fold in one round of hybridization reaction (Albert et al., 2007). Multiple enrichment cycles have been used to further increase target enrichment efficiency (Lee et al., 2009; Summerer et al., 2009).

Both liquid-phase and solid-phase hybridization methods have successfully been used for enrichment of genes involved in deafness (Brownstein et al., 2011; Shearer et al., 2010; Tang et al., 2011). One key factor for a successful target enrichment outcome is the probe design, which is usually based on information from gene annotation databases (e.g., consensus coding sequence (CCDS) Database, RefSeq Database, etc.). Therefore, unknown disease gene exons, evolutionarily conserved non-coding genome regions, and regulatory sequences (e.g., enhancers or promoters) are typically not included in the designs. Regions containing a high GC percentage may also hybridize poorly for enrichment (Okou et al., 2007).

Non-hybridization-based enrichment methods

Other PCR-based non-hybridization gene enrichment schemes include Molecular Inversion Probes (MIP) (Deng et al., 2009; Porreca et al., 2007; Turner et al., 2009) and Spacer Multiplex Amplification Reaction (SMART) methods (Krishnakumar et al., 2008). However, neither of these methods was developed into commercially available products. Chromosome sorting (Ibrahim et al., 2004) can be utilized if either large sections of the chromosome or the complete chromosome is the target. Flow-sorted chromosomes can be sequenced by standard NGS protocols after DNA fragmentation. Sequencing of manually sorted metaphase chromosomes by microdissection has also been reported (Weise et al., 2010). These nonhybridization-based methods require specialized instruments and are not suited for high-throughput applications and consequently are most likely applicable only to unique research cases.

Highly multiplexed PCRs have inherent limitations, and usually a high percentage of targeted regions fail to amplify. Another disadvantage is that the enrichment results for particular genes in the mix are impossible to verify before the sequencing results are available. Recent innovative applications of PCR in microfluidic environments have helped overcome these weaknesses. Oil microdroplets segregate thousands of individual PCR reactions in the same reaction tube. Each microdroplet is generated in a way which contains a simple combination of primer pairs and genomic DNA mix, allowing for a more reliable and successful multiplexed PCR outcome (Tewhey et al., 2009). However, the initial cost of purchasing a large amount of PCR primers and the cost of equipment are very substantial in commercial offerings (e.g., RainDance Inc, Lexington, MA).

Another commercially available approach, offered by Halo Genomics, uses restriction enzymes to fragment DNA first in order to obtain DNA fragments with known end sequences (Dahl et al., 2005). Probes specific to the targeted fragments are hybridized and PCR reactions are designed to make all targeted DNA fragments circular. The circular DNAs are separated by biotinylated probes which react with streptavidin-affiliated magnetic beads in order to achieve the purification of targeted genomic regions.

Data analysis for NGS applications

With the availability of human genome and exome sequencing results from many individuals, it is now clear that two unrelated individuals have at least two million differences in their genomic DNA sequences (Moore et al., 2011). Such a large variability in

normal healthy individuals makes the designation of a “reference human genome” obsolete. While no one individual’s genome can be readily named a “reference,” the National Center for Biotechnology Information (NCBI) and many other databases maintain and update a standard assembly of the human genome that is derived from many individuals (e.g., HG18, HG19). Any deviation from the standard reference genome maintained by NCBI will be referred to as a variant. When the variant has clinical consequences, it is labeled as a pathogenic variant. When the differences in DNA are benign variants, they are identified as SNPs or MNPs. The variants may also be designated as variations with unknown significance if the effect of the DNA change is unknown. A more comprehensive classification of variants has been proposed recently (Richards et al., 2008).

Strategies for analyzing NGS reads

The first step in finding DNA variations is to align the NGS reads (short DNA segments, usually with a length of 25 to 100 bps) with the reference genome (Ruffalo et al., 2011). Recently a unified analytic framework to identify genetic variations among multiple samples has been proposed (DePristo et al., 2011). For researchers with bioinformatic training and knowledge of the Linux system, many open source aligners are available. A summary about the sources for downloading various software packages is given at <http://seqanswers.com/forums/showthread.php?t=43>. Many integrated commercial solutions (e.g., Softgenetics NextGene) are available for data analysis, and some of them are web-based cloud-computational servers (e.g., www.dnanexus.com). Nucleotides in the regions of interest are usually covered many times because of the random nature of the short NGS reads. The depth of coverage is defined as the number of times each nucleotide is independently sequenced in difference reads, which is a unique idea in NGS data analysis when compared to the Sanger sequencing method (Bao et al., 2011). After analysis, differences between the subject’s sequence and the reference sequence are reported in a list that usually contains large numbers of variants. The next step is to determine the significance of these variants by sorting out benign SNPs and those that cause diseases.

Strategies to identify pathogenic mutations using the NGS approach generally rely on certain assumptions. It is more likely that the mutations having a higher penetrance and those affecting the protein structure with significant functional consequences will be identified. Candidates for such mutations are generally non-synonymous SNPs/MNPs, insertions/deletions, and splice-site mutations. A growing number of databases and software packages (Table 1) are used to determine the meaning of variants discovered by NGS (Bao et al., 2011; Galperin et al., 2011). Sorting out large numbers of variants to find pathogenic mutations is still a challenging task.

Systematic sequencing and misalignment errors have to be removed first in alignment algorithms. The increasing amount of knowledge in many online databases (Table 1) such as the Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov/omim>) and the Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>) will certainly help with the interpretation of NGS results. A combined approach including multistep filtering by using phenotype databases (Table 1) in consideration of the patient’s phenotype (2010; Altshuler et al., 2010), patterns of inheritance among family members, and the computational approach to predict potential effects of the variant on the protein function (e.g., SIFT analysis) (Ng et al., 2003) are usually required to identify candidates of deleterious mutations. With improving genotype and phenotype definitions over time, it would be prudent to review subjects’ NGS results regularly for new information when no definitive conclusion is initially achieved.

In clinical applications, it seems likely that targeted sequencing will be useful alongside WES and WGS for a long time due to the much lower demands on bioinformatics personnel,

infrastructure, and turnaround time. The amount of sequencing needed for a panel of genes associated with deafness (supplemental Table 1) is about four orders of magnitude less than WGS (0.3 million bps vs. 3.2 billion bps). Although targeted capture data may also contain a large number of variants, the amount will be orders of magnitude lower, making secondary analyses and data management more straightforward and less time-consuming. A typical NGS mutation report should contain basic information such as the gene name and gene description, chromosomal position of variants, zygosity of the variant, type of nucleotide change, predicted protein change, the sequencing quality of each variant, etc. Further items in the report may include web links to entries in the Online Mendelian Inheritance in Man (OMIM), scale-invariant feature transform (SIFT) (Ng et al., 2003) prediction of how likely the change is to damage protein function, interspecies conservation of each nucleotide, dbSNP entry for the found variants, and allele frequency from the 1,000 Genomes Project (Altshuler et al., 2010). The report also needs to list the genes/exons that happen to have poor quality reads. The regions that were not or could not be targeted (e.g., due to high GC contents or repetitive regions) by the capture baits should also be mentioned. These regions should be Sanger sequenced where possible to gain a more complete picture of the targeted genes. After a targeted NGS scan, it is important to know the name and proportion of genes for which the test is not able to make a confident determination, due to either poor target enrichment or poor quality of sequencing data. In the United States, results generated by the NGS approach are currently not CLIA-approved. Therefore, it is essential for the mutation results of clinical significance to be verified by Sanger sequencing.

Common sources of errors in NGS data

Both systematic false positive and false negative variation calls exist in NGS results. False positives most often arise from incorrect genome mapping, systematic equipment sequencing errors, or sequencing calls of DNA fragments near the end of equipment capability. For example, NGS systems based on the pyrosequencing principle are known to have systematic errors when reading a strain of more than five or six of the same nucleotides, known as homopolymer length inaccuracies (Margulies et al., 2005). Systematic equipment errors occur in many unrelated samples and can be removed from the final list of variants. Artifacts due to errors in software alignment may be reduced by cross examination with different analysis packages. False negative results are usually caused by a low depth of coverage in particular targeted regions, poor capture efficiency, and difficulty in unambiguously aligning repetitive regions (Nothnagel et al., 2011). Although NGS technologies generally have higher individual-base calling error rates than Sanger sequencing in a single read, the NGS error rate is significantly reduced by increasing the depth of sequencing coverage (Koboldt et al., 2010). However, Sanger sequencing is usually required before the final report can be issued for clinical purposes, because results obtained by NGS approaches are not currently CLIA-certified.

Increasing coverage effectively reduces the error rate, especially for sequencing calls of heterozygous variants. Longer reads are generally better than shorter ones, as they reduce false positives from mapping ambiguity. When longer reads are used, the nucleotide calls near the end can be trimmed, as they tend to have much higher error rates (Ledergerber et al., 2011). Paired-end reads add one more dimension of constraint for alignment: the approximate distance between the two reads. Such a constraint helps to achieve better alignments that are not possible with single-end reads. Paired-end reads greatly help alignment software to find the correct target region and to reduce the interference from pseudogene regions (Nielsen et al., 2011).

Application of targeted NGS in studying human diseases

Most Mendelian disorders are caused by exonic or splice-site mutations that alter the amino acid sequence of the affected gene. An exome constitutes only ~1% of the human genome; however, ~85% of disease-related mutations found so far are in the protein coding regions (Teer et al., 2010). The number of known mutations in human somatic genes underlying or associated with inherited disease exceeds 110,000 in more than 3,700 different genes (e.g., entries found in the Human Gene Mutation Database). In addition, the entire human gene pool also undergoes spontaneous mutations resulting in new pathogenic alleles (Davies et al., 2010).

In classical strategies of identifying disease-associated mutations, linkage analysis in family members with shared genetic relations is performed. With a successful outcome, candidate genomic regions enclosing the disease gene may be narrowed down to a specific region. Systematic sequencing of whole regions or implementing a candidate gene approach using known biological knowledge may locate the genes within a specific chromosome interval. NGS has brought new ways of addressing monogenic disorders. Because of its large capacity to unbiasedly survey the whole exome and whole genome, NGS is well suited in the usage of discovering the cause of rare genetic disorders. It is estimated that causative mutations are known for about half of the genetic diseases listed in the database of Mendelian disorders in humans (<http://www.ncbi.nlm.nih.gov/omim>) (Bick et al., 2011). Many of the diseases are rare, which may create difficulties for linkage analysis in which data collections from large families are often needed (Bick et al., 2011).

When properly executed, the WES approach may dramatically reduce the required sample numbers needed for a successful outcome. The use of NGS has frequently resulted in identifying disease genes with even a limited number of patient samples (Kalay et al., 2011; Krawitz et al., 2010; Kuhlenbaumer et al., 2011; Musunuru et al., 2010; Ng et al., 2010b; Ng et al., 2010c; Puente et al., 2011; Simpson et al., 2011). Novel and causative variants have recently been discovered for diverse types of diseases, including neuropathy cases (Brkanac et al., 2009), Clericuzio type poikiloderma with neutropenia (Volpi et al., 2010), familial exudative vitreoretinopathy (Nikopoulos et al., 2010), immunological disorders (Bolze et al., 2010; Byun et al., 2010), intellectual disabilities (Abou Jamra et al., 2011; Shoubbridge et al., 2010), cancer predisposition (Shoubbridge et al., 2010), and other abnormalities (Barak et al., 2011; Bilguvar et al., 2010; Otto et al., 2010).

The NGS approach has identified disease-causing genetic mutations from either unrelated subjects (Choi et al., 2009; Gilissen et al., 2010; Hoischen et al., 2010; Ng et al., 2009; Ng et al., 2010c; Otto et al., 2010) or from family members (Anastasio et al., 2010; Erlich et al., 2011; Glazov et al., 2011; Krawitz et al., 2010; Musunuru et al., 2010; Ng et al., 2010b; Tsurusaki et al., 2011). WES data from members of the same family offer the advantage of cross checking for benign SNPs, therefore hastening the mutation discoveries. WES was also successful in identifying causative mutations in diseases with genetic and phenotypic heterogeneity (Isidor et al., 2011; Johnson et al., 2010; Simpson et al., 2011; Wang et al., 2010), which is difficult for traditional linkage-based approaches.

Novel genes for non-syndromic (Rehman et al., 2010; Walsh et al., 2010) and syndromic (Pierce et al., 2010; Zheng et al., 2011) hearing loss were also identified recently by the targeted NGS approach. These studies show that the targeted genomic region (Rehman et al., 2010) or whole exome NGS (Pierce et al., 2010; Walsh et al., 2010), followed by verification from nonconsanguineous families, and by functional and immunolabeling examinations, can reveal critical disease-causing genes from small pedigrees. Rehman captured a targeted genomic region at the DFNB79 locus on chromosome 9q34.3 (including

108 genes) and followed with NGS of the 2.9 Mbp region. One affected member of a consanguineous family was initially tested. Homozygosity for a total of eight previously unreported variants was detected in the targeted region. Six of them were determined to be polymorphisms and one was in a noncoding region. The only remaining variant was a nonsense mutation in a predicted gene, *C9orf75* (later renamed as *TPRN*). Follow-up tests in affected subjects with DFNB79 found four truncating alleles of this gene, and immunolocalization of the *TPRN* protein product in the mouse cochlea showed prominent expression in the taper region of hair cell stereocilia.

In the Walsh et al. study, the whole exome was sequenced, and variations at the DFNB82 locus were analyzed. After filtering out polymorphisms from the whole exome sequence by using publicly available and population-specific databases, only a single deleterious homozygous mutation remained. They followed with functional and immunolabeling studies, and evidence further supported the theory that mutations resulting in an early truncation of the G protein-signaling modulator *GPSM2* are the cause of DFNB82. In the Pierce et al. study, the authors used gDNA from two sisters in a family diagnosed with well-characterized Perrault syndrome. WES revealed exactly one gene (*HSD17B4*) with two rare functional variants. Both sisters are compound heterozygotes for *HSD17B4* c.650A>G (p.Y217C) and *HSB17B4* c.1704T>A (p.Y568X). The mutations are predicted by structural analysis to destabilize the *HSD17B4* dehydrogenase domain. They also found that protein expression of mutant *HSD17B4* in a compound heterozygote was severely reduced.

NGS technologies also enabled classification of many types of somatically acquired mutations in cancers (Pleasance et al., 2010; Stratton et al., 2009). Sequencing specific oncogenes and/or tumor suppressor genes at very high coverage for heterogeneous samples with a small percentage of tumor cells was made feasible by NGS approaches (Thomas et al., 2006). WES has been used to identify biomarkers in individuals with acute myeloid leukemia (Ley et al., 2010; Mardis et al., 2009; Yan et al., 2011). Other innovative uses of WES are for designing personalized chemotherapy (Wesolowska et al., 2011) and for classification of prognostic outcomes of chronic myelomonocytic leukemia based on patterns of mutations (Kohlmann et al., 2010).

Most of the identified variants thus far were either small deletions or non-synonymous substitutions, and were found in the exonic regions. However, splice-site mutations that disrupt a translation resulting in exon skip and a frame shift were also found (Volpi et al., 2010). One of the unique strengths of WGS is that it can be used to identify the breakpoints in balanced chromosome translocations and inversions (Talkowski et al., 2011). This permits the identification of genes linked to the phenotype that results from *de novo* chromosomal rearrangements.

Current clinical applications of targeted NGS for a focused panel of disease genes

The success of NGS in research has already resulted in its translational uses in clinical care, and many of them are for diagnostic mutation detection of focused panels of disease genes. For example, clinical diagnosis of a panel of >90 X-Linked intellectual disability genes (<http://genetics.emory.edu/egl/tests/testpage.php?testid=1111>) and a pan-cardiomyopathy panel of 46 genes (<http://pcpgm.partners.org/imm/tests/cardiomyopathy>) are offered now. NGS testing of up to 84 human genes implicated in both syndromic and non-syndromic hearing loss is also offered on the market (http://www.healthcare.uiowa.edu/labs/morl/index_CDS.htm and www.otogenetics.com).

In a recent case of clinical practice, NGS was successfully used in finding the causative mutations for a child with a severe Crohn's disease-like illness (Worthey et al., 2011). Analysis of the patient's WES results revealed a mutation in the *XIAP* gene. This finding led

to the selection of an effective treatment by hemopoietic progenitor cell transplant (Worthey et al., 2011). In another successful WES clinical application case, a clinical decision was made for liver transplantation for an infant with acute liver failure (Goh et al., 2011).

In contrast to complicated genetic traits affecting the incidence of cancer, heart disease, and neurological disorders, deafness caused by genetic mutations is monogenic in most cases (Hilgert et al., 2009). The deafness phenotypes are usually predicted by genetic testing with high accuracy (Robin et al., 2005). Most newborn hearing-loss cases are caused by genetic mutations, especially in developed countries (Smith et al., 2005). Therefore, performing genetic tests prior to other testing may limit or eliminate the necessity for conducting other expensive tests, as a definitive diagnosis can often be made based on genetic testing alone. NGS is a promising technology for such a genetic test with a capacity of analyzing all known and candidate genes involved in deafness for routine clinical applications in one test. As we have pointed out earlier, mutation results generated by the NGS approach will require Sanger verification for clinical test reports, as NGS is not currently a CLIA-approved approach.

Shearer et al. (Shearer et al., 2010) assessed the feasibility of targeted enrichment and NGS for 54 genes known to cause non-syndromic sensorineural hearing loss. Both solid-phase and solution-based methods were tested for targeted gene capture. Pathogenic mutations were identified using a custom-variant calling platform that incorporates the analysis of protein-damaging effects by publicly available pathogenicity prediction tools (e.g., SIFT, BLOSUM, Polyphen2, and Align-GVGD). Among the six idiopathic hearing-loss patient samples, the authors identified the pathogenic mutations in five of them. They concluded that targeted NGS technologies provide “sensitivity, specificity, and reproducibility at levels sufficient to perform genetic diagnosis of hearing loss.”

In another recent NGS application of identifying gene mutations in deafness, Brownstein et al. (Brownstein et al., 2011) used a custom-designed set of oligonucleotide probes that targeted 85 human and 161 mouse genes known to cause deafness when mutated. Multiple samples were barcoded and sequenced in one lane using the Illumina NGS platform by a paired-end sequencing protocol. The median coverage of all the exons ranged between 757 and 2,080, and 95% of the targeted bases were covered by at least 10 independent reads. The study identified disease-causing mutations in *CDH23*, *MYO15A*, *TECTA*, *TMC1*, and *WFS1* genes from six of the 11 original probands and their families, and helped the identification of causative alleles in additional family members. Brownstein et al. (2011) found a founder allele mutation in the *TMC1* for the Moroccan Jewish population. For deaf Israelis of Moroccan Jewish ancestry, a single recessive mutation in *TMC1* (p.S647P) was proved to be a founder allele, contributing to 34% of genetic hearing loss in this population.

These recent studies demonstrate the feasibility of conducting diagnostic tests for all known genes involved in deafness by the NGS approach (Brownstein et al., 2011; Shearer et al., 2010). However, the high per-sample cost for capturing targeted exons is still a major barrier for the widespread use of NGS technologies on a population level. We have explored the use of a cDNA-based probe for targeted gene enrichment in order to reduce costs (Tang et al., 2011). In-house production of cDNA-based bait probes were used for capturing exons of about 84 genes. When 46 samples were pooled in one HiSeq2000 lane, the average coverage achieved was 263 ± 12 ($n=46$). The majority of the targeted 1,253 exons ($86.4 \pm 3.0\%$, $n=46$) are covered more than 100 times. It appears that the cDNA-based bait probe is a cost-effective alternative to commercially available sources for producing hybridization probes needed in targeted gene enrichment (Tang et al., 2011).

Current limitations and regulatory issues of NGS technologies

NGS is still in its infant stage for clinical applications, and a lack of clear regulatory oversight over NGS data quality control, data analysis standards, and standardization of data reporting may hinder its rapid translation (Bunnik et al., 2011). Some recent attempts have been initiated in order to establish laboratory procedures and regulations for quality control and validation using alternative assessments. For example, the Centers for Disease Control and Prevention sponsored a conference in 2011 on Next-Generation Sequencing, Standardization of Clinical Testing. There are still many NGS data analysis issues remaining to be solved. How to standardize and compare data across different NGS platforms when the same samples are tested is still an unresolved issue (Nothnagel et al., 2011). One of the known weaknesses of the NGS application is its use in determining the number of trinucleotide repeats that are the pathogenic basis for many forms of intellectual disability (e.g., FMR1). Error rates of NGS data for these applications were higher because the base elongation at the repeat regions during sequencing was unreliable (Nakamura et al., 2011).

In addition, there are controversial ethical issues surrounding genome research and its applications. The use and storage of patient information and data are stringently regulated by institutional review boards (IRBs) in the research setting. However, specific rules governing the use of NGS data are currently lacking. Definitive personal identifier information could be deduced from the large amount of information contained in the NGS data, not just for the patient involved but also for the patient's relatives and related ethnic groups. Although personal labels are usually removed during data analysis, sensitive personal information could still be identified due to the large amount of genomic information contained in the WES and WGS data. The NGS data are often stored on computer servers and shared among collaborative research groups. Sometimes online vendors are used for this purpose. Many research funding sources demand a rapid sharing of NGS data and often require the depositing of data sets in open-access archives as a condition for publishing papers (Kaye et al., 2009). The complex issues in handling genomic data in regard to data storage policy and the specifications for deposit into publicly accessible databases make meaningful informed consent to WGS and WES projects difficult to define. Reporting research findings back to the participants has been considered by some as an important part of maintaining public trust in tax-dollar funded research (Caulfield et al., 2008; Lacroix et al., 2008). However, informing individual participants before results are published continues to be controversial, particularly for WGS and WES studies.

Another important concern for laboratories that carry out large-scale resequencing projects and diagnosis using WES, WGS, and focused-panel NGS approaches is the gene sequence patents issued by United States Patent and Trademark Office. A recent case brought by the Association for Molecular Pathology has questioned the validity of such patents on gene sequences. A preliminary ruling has found that the gene patent in question was invalid, and the case has been appealed (Cook-Deegan et al., 2010).

Conclusions and future directions

Advances in NGS technologies are accelerating the discovery of genetic disorders at an unprecedented pace. Vast amounts of WGS, WES, and focused-panel NGS data are now being generated by laboratories worldwide on a scale that was unimaginable just a few years ago. Applications of NGS technologies are now beginning to enter clinical practice. Interpreting the data and translating the research results into applications that improve healthcare is still challenging. Sifting through the millions of variants in an individual's genome for the pathogenic mutation seems to be the most urgent task at hand. The creation of dedicated databases specifically for the purpose of clinical interpretation based on NGS results from a large number of normal controls and diagnosed patients will significantly help

this endeavor. Another important aspect is the concurrent development of genetic counseling capabilities to interpret the large amount of data revealed by NGS for clinical use. In the near future, physicians may combine a past medical history and family history with NGS diagnostic data to identify disease predisposition variants and variants that affect drug metabolism in individuals. Carrier status for genetic disorders could also be a part of the patients' medical history. Improved applications of NGS may help define genetic profiles of patients and accelerate the pace of personalized medical care.

Genetic screening of a panel of genes involved in deafness is one of the potential applications of NGS. Universal Newborn Hearing Screening (UNHS) is standard clinical practice protocol in the United States, Canada, Europe and many developing countries (Morton et al., 2006), now covering about 93% of newborns in the United States (White, 2003). Referral rates range from 0.8% to 0.3% (Clemens et al., 2001). In contrast, the rate of confirmed congenital deafness is estimated to be 0.1% (Smith et al., 2005). In addition to high false-positive rates, other shortcomings of the current physiology-based test methods for UNHS include (1) the inability to determine the underlying cause of hearing loss and (2) the inability to detect early-onset childhood hearing loss, which affects more patients than congenital defects. According to CDC data (<http://www.cdc.gov/ncbddd/ehdi/data.htm>), hearing loss prevalence increases from 1‰ at birth to 6‰ for school-age children. A third reason is necessity of one's physical presence for repeated examinations to confirm the accuracy of the initial screening result. About 60% of the patients who failed the first round of physiology-based tests never come back for a follow-up test (<http://www.cdc.gov/ncbddd/ehdi/>). Most congenital and many early childhood deafness cases are caused by genetic mutations (Smith et al., 2005). Diagnostic genetic information of individual genes involved in deafness has already been used successfully in patient treatment, management, and genetic counseling. For example, patients with A1555G mitochondrial mutations can be warned about their increased susceptibility for aminoglycoside ototoxicity. With the development of appropriate gene enrichment and NGS technologies, all known and candidate genes involved in deafness can be examined in one test in order to give physicians a more complete survey of the genetic basis underlying a patient's hearing loss. Genetic screening of multiple genes involved in deafness based on NGS technologies may therefore significantly enhance the care of people who are affected by hearing loss due to genetic causes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by grants to Lin from the National Institute on Deafness and other Communication Disorders (NIDCD 4R33DC010476, 1R41DC009713 and RO1 DC006483). Tang received grant support from NIDCD (R21 DC008672). This research project was supported in part by the GRA genome center of the Emory University School of Medicine. We thank Ms. Anne Lin for proofreading the final version of the manuscript.

References

- Human genome: Genomes by the thousand. *Nature*. 2010; 467:1026–7. [PubMed: 20981067]
- Abou Jamra R, Philippe O, Raas-Rothschild A, Eck SH, Graf E, Buchert R, Borck G, Ekici A, Brockschmidt FF, Nothen MM, Munnich A, Strom TM, Reis A, Colleaux L. Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet*. 2011; 88:788–95. [PubMed: 21620353]

- Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. *Nat Methods*. 2007; 4:903–905. [PubMed: 17934467]
- Altshuler D, Durbin R, Abecasis G, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–73. [PubMed: 20981092]
- Anastasio N, Ben-Omran T, Teebi A, Ha KC, Lalonde E, Ali R, Almureikhi M, Der Kaloustian VM, Liu J, Rosenblatt DS, Majewski J, Jerome-Majewska LA. Mutations in SCARF2 are responsible for Van Den Ende-Gupta syndrome. *Am J Hum Genet*. 2010; 87:553–9. [PubMed: 20887961]
- Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch MJ, D'Ascenzo M, Kitzman J, Wu YQ, Newsham I, Richmond TA, Jeddloh JA, Muzny D, Albert TJ, Gibbs RA. Whole exome capture in solution with 3 Gbp of data. *Genome biology*. 2010; 11:R62. [PubMed: 20565776]
- Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ. Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet*. 2011; 56:406–14. [PubMed: 21525877]
- Barak T, Kwan KY, Louvi A, Demirbilek V, Saygi S, Tuysuz B, Choi M, Boyaci H, Doerschner K, Zhu Y, Kaymakcalan H, Yilmaz S, Bakircioglu M, Caglayan AO, Ozturk AK, Yasuno K, Brunken WJ, Atalar E, Yalcinkaya C, Dincer A, Bronen RA, Mane S, Ozelik T, Lifton RP, Sestan N, Bilguvar K, Gunel M. Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nature genetics*. 2011; 43:590–4. [PubMed: 21572413]
- Bick D, Dimmock D. Whole exome and whole genome sequencing. *Curr Opin Pediatr*. 2011
- Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tuysuz B, Caglayan AO, Gokben S, Kaymakcalan H, Barak T, Bakircioglu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yilmaz S, Dincer A, Johnson MH, Bronen RA, Kocer N, Per H, Mane S, Pamir MN, Yalcinkaya C, Kumandas S, Topcu M, Ozmen M, Sestan N, Lifton RP, State MW, Gunel M. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*. 2010; 467:207–10. [PubMed: 20729831]
- Bolze A, Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL. Whole-exome-sequencing-based discovery of human FADD deficiency. *Am J Hum Genet*. 2010; 87:873–81. [PubMed: 21109225]
- Borgstrom E, Lundin S, Lundberg J. Large scale library generation for high throughput sequencing. *PLoS ONE*. 2011; 6:e19119. [PubMed: 21589638]
- Brkanac Z, Spencer D, Shendure J, Robertson PD, Matsushita M, Vu T, Bird TD, Olson MV, Raskind WH. IFRD1 is a candidate gene for SMNA on chromosome 7q22-q23. *Am J Hum Genet*. 2009; 84:692–7. [PubMed: 19409521]
- Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, Rayyan AA, Parzefall T, Lev D, Shalev S, Frydman M, Davidov B, Shohat M, Rahile M, Lieberman S, Levy-Lahad E, Lee MK, Shomron N, King MC, Walsh T, Kanaan M, Avraham KB. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families. *Genome biology*. 2011; 12:R89. [PubMed: 21917145]
- Bunnik EM, Schermer MH, Janssens AC. Personal genome testing: test characteristics to clarify the discourse on ethical, legal and societal issues. *BMC Med Ethics*. 2011; 12:11. [PubMed: 21672210]
- Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao C, Jouanguy E, Feske S, Abel L, Casanova JL. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J Exp Med*. 2010; 207:2307–12. [PubMed: 20876309]
- Caulfield T, McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA, Juengst ET, Kaye J, Kedes L, Knoppers BM, Lemmens T, Meslin EM, Murphy J, Nussbaum RL, Otlowski M, Pullman D, Ray PN, Sugarman J, Timmons M. Research ethics recommendations for whole-genome research: Consensus statement. *PLoS Biol*. 2008; 6:430–435.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:19096–101. [PubMed: 19861545]

- Clark MJ, Chen R, Lam HY, Karczewski KJ, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology*. 2011; 29:908–14.
- Clemens CJ, Davis SA. Minimizing false-positives in universal newborn hearing screening: a simple solution. *Pediatrics*. 2001; 107:E29. [PubMed: 11230610]
- Cook-Deegan R, Heaney C. Patents in genomics and human genetics. *Annu Rev Genomics Hum Genet*. 2010; 11:383–425. [PubMed: 20590431]
- Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods*. 2008; 5:887–93. [PubMed: 18794863]
- Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic acids research*. 2005; 33:e71. [PubMed: 15860768]
- Davies MA, Samuels Y. Analysis of the genome to personalize therapy for melanoma. *Oncogene*. 2010; 29:5545–55. [PubMed: 20697348]
- Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature biotechnology*. 2009; 27:353–60.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011; 43:491–8. [PubMed: 21478889]
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009; 323:133–8. [PubMed: 19023044]
- Erich Y, Edvardson S, Hodges E, Zenvirt S, Thekkat P, Shaag A, Dor T, Hannon GJ, Elpeleg O. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome research*. 2011; 21:658–64. [PubMed: 21487076]
- Galperin MY, Cochrane GR. The 2011 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic acids research*. 2011; 39:D1–6. [PubMed: 21177655]
- Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet*. 2010; 87:418–23. [PubMed: 20817137]
- Glazov EA, Zankl A, Donskoi M, Kenna TJ, Thomas GP, Clark GR, Duncan EL, Brown MA. Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet*. 2011; 7:e1002027. [PubMed: 21455487]
- Goh V, Helbling D, Biank V, Jarzembowski J, Dimmock D. Next Generation Sequencing Facilitates The Diagnosis In A Child With Twinkle Mutations Causing Cholestatic Liver Failure. *J Pediatr Gastroenterol Nutr*. 2011
- Herman DS, Hovingh GK, Iartchouk O, Rehm HL, Kucherlapati R, Seidman JG, Seidman CE. Filter-based hybridization capture of subgenomes enables resequencing and copy-number detection. *Nat Methods*. 2009; 6:507–10. [PubMed: 19543287]
- Hilgert N, Smith RJ, Van Camp G. Forty-six genes causing nonsyndromic hearing impairment: Which ones should be analyzed in DNA diagnostics? *Mutat Res*. 2009; 681:189–96. [PubMed: 18804553]
- Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. *Nature genetics*. 2007; 39:1522–7. [PubMed: 17982454]

- Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BBA, Veltman JA. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature genetics*. 2010; 42:483–485. [PubMed: 20436468]
- Ibrahim SF, van den Engh G. High-speed chromosome sorting. *Chromosome Res*. 2004; 12:5–14. [PubMed: 14984097]
- Inoue T, Orgel LE. A nonenzymatic RNA polymerase model. *Science*. 1983; 219:859–62. [PubMed: 6186026]
- Isidor B, Lindenbaum P, Pichon O, Bezieau S, Dina C, Jacquemont S, Martin-Coignard D, Thauvin-Robinet C, Le Merrer M, Mandel JL, David A, Faivre L, Cormier-Daire V, Redon R, Le Caignec C. Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nature genetics*. 2011; 43:306–8. [PubMed: 21378989]
- Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurro MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G, Galassi G, Scholz SW, Taylor JP, Restagno G, Chio A, Traynor BJ. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron*. 2010; 68:857–64. [PubMed: 21145000]
- Kalay E, Yigit G, Aslan Y, Brown KE, Pohl E, Bicknell LS, Kayserili H, Li Y, Tuysuz B, Nurnberg G, Kiess W, Koegl M, Baessmann I, Buruk K, Toraman B, Kayipmaz S, Kul S, Ikbil M, Turner DJ, Taylor MS, Aerts J, Scott C, Milstein K, Dollfus H, Wieczorek D, Brunner HG, Hurles M, Jackson AP, Rauch A, Nurnberg P, Karaguzel A, Wollnik B. CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nature genetics*. 2011; 43:23–6. [PubMed: 21131973]
- Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. SCIENCE AND SOCIETY Data sharing in genomics - re-shaping scientific practice. *Nat. Rev. Genet*. 2009; 10:331–335. [PubMed: 19308065]
- Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform*. 2010; 11:484–98. [PubMed: 20519329]
- Kohlmann A, Grossmann V, Klein HU, Schindela S, Weiss T, Kazak B, Dicker F, Schnittger S, Dugas M, Kern W, Haferlach C, Haferlach T. Next-Generation Sequencing Technology Reveals a Characteristic Pattern of Molecular Mutations in 72.8% of Chronic Myelomonocytic Leukemia by Detecting Frequent Alterations in TET2, CBL, RAS, and RUNX1. *J. Clin. Oncol*. 2010; 28:3858–3865. [PubMed: 20644105]
- Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Kohler S, Jager M, Grunhagen J, de Condor BJ, Doelken S, Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nature genetics*. 2010; 42:827–9. [PubMed: 20802478]
- Krishnakumar S, Zheng J, Wilhelmy J, Faham M, Mindrinos M, Davis R. A comprehensive assay for targeted multiplex amplification of human DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:9296–301. [PubMed: 18599465]
- Kuhlenbaumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat*. 2011; 32:144–51. [PubMed: 21280146]
- Lacroix M, Nycum G, Godard B, Knoppers BM. Should physicians warn patients' relatives of genetic risks? *Can. Med. Assoc. J*. 2008; 178:593–595. [PubMed: 18299550]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston

- RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]
- Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. *Brief Bioinform*. 2011; 12:489–97. [PubMed: 21245079]
- Lee H, O'Connor BD, Merriman B, Funari VA, Homer N, Chen Z, Cohn DH, Nelson SF. Improving the efficiency of genomic loci capture using oligonucleotide arrays for high throughput resequencing. *BMC genomics*. 2009; 10:646. [PubMed: 20043857]
- Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang QY, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med*. 2010; 363:2424–2433. [PubMed: 21067377]
- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J. Med. Genet*. 2011; 48:580–589. [PubMed: 21730106]
- Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipiek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 2009; 361:1058–66. [PubMed: 19657110]
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437:376–80. [PubMed: 16056220]
- McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, et al. A physical map of the human genome. *Nature*. 2001; 409:934–41. [PubMed: 11237014]
- Moore B, Hu H, Singleton M, Reese MG, De La Vega FM, Yandell M. Global analysis of disease-related DNA sequence variation in 10 healthy individuals: implications for whole genome-based clinical diagnostics. *Genet Med*. 2011; 13:210–7. [PubMed: 21325948]

- Morton CC, Nance WE. Newborn hearing screening--a silent revolution. *N Engl J Med*. 2006; 354:2151–64. [PubMed: 16707752]
- Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernysky A, Gonzalez E, Rudzicz N, Engert JC, Depristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S. Exome Sequencing, ANGPTL3 Mutations, and Familial Combined Hypolipidemia. *N Engl J Med*. 2010
- Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S. Sequence-specific error profile of Illumina sequencers. *Nucleic acids research*. 2011; 39:e90. [PubMed: 21576222]
- Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic acids research*. 2003; 31:3812–3814. [PubMed: 12824425]
- Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Human molecular genetics*. 2010a; 19:R119–24. [PubMed: 20846941]
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*. 2010b; 42:30–5. [PubMed: 19915526]
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009; 461:272–6. [PubMed: 19684571]
- Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature genetics*. 2010c; 42:790–3. [PubMed: 20711175]
- Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nature reviews*. 2011; 12:443–51.
- Nikopoulos K, Gilissen C, Hoischen A, van Nouhuys CE, Boonstra FN, Blokland EA, Arts P, Wieskamp N, Strom TM, Ayuso C, Tilanus MA, Bouwhuis S, Mukhopadhyay A, Scheffer H, Hoefsloot LH, Veltman JA, Cremers FP, Collin RW. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet*. 2010; 86:240–7. [PubMed: 20159111]
- Nothnagel M, Herrmann A, Wolf A, Schreiber S, Platzer M, Siebert R, Krawczak M, Hampe J. Technology-specific error signatures in the 1000 Genomes Project data. *Hum Genet*. 2011
- Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods*. 2007; 4:907–909. [PubMed: 17934469]
- Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, Patil SB, Levy S, Ghosh AK, Murga-Zamalloa CA, van Reeuwijk J, Letteboer SJ, Sang L, Giles RH, Liu Q, Coene KL, Estrada-Cuzcano A, Collin RW, McLaughlin HM, Held S, Kasanuki JM, Ramaswami G, Conte J, Lopez I, Washburn J, Macdonald J, Hu J, Yamashita Y, Maher ER, Guay-Woodford LM, Neumann HP, Obermuller N, Koenekoop RK, Bergmann C, Bei X, Lewis RA, Katsanis N, Lopes V, Williams DS, Lyons RH, Dang CV, Brito DA, Dias MB, Zhang X, Cavalcoli JD, Nurnberg G, Nurnberg P, Pierce EA, Jackson PK, Antignac C, Saunier S, Roepman R, Dollfus H, Khanna H, Hildebrandt F. Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nature genetics*. 2010; 42:840–50. [PubMed: 20835237]
- Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet*. 2010; 87:282–8. [PubMed: 20673864]
- Pleasant ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordonez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen LN, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia MM, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning ZM, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L,

- Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2010; 463:191–U73. [PubMed: 20016485]
- Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J. Multiplex amplification of large sets of human exons. *Nat Methods*. 2007; 4:931–6. [PubMed: 17934468]
- Puente XS, Quesada V, Osorio FG, Cabanillas R, Cadinanos J, Fraile JM, Ordonez GR, Puente DA, Gutierrez-Fernandez A, Fanjul-Fernandez M, Levy N, Freije JM, Lopez-Otin C. Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. *Am J Hum Genet*. 2011; 88:650–6. [PubMed: 21549337]
- Rehman AU, Morell RJ, Belyantseva IA, Khan SY, Boger ET, Shahzad M, Ahmed ZM, Riazuddin S, Khan SN, Friedman TB. Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. *Am J Hum Genet*. 2010; 86:378–88. [PubMed: 20170899]
- Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008; 10:294–300. [PubMed: 18414213]
- Robin NH, Prucka SK, Woolley AL, Smith RJ. The use of genetic testing in the evaluation of hearing impairment in a child. *Curr Opin Pediatr*. 2005; 17:709–12. [PubMed: 16282775]
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011; 475:348–52. [PubMed: 21776081]
- Ruffalo M, LaFramboise T, Koyuturk M. Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics (Oxford, England)*. 2011; 27:2790–6.
- Ruparel H, Bi L, Li Z, Bai X, Kim DH, Turro NJ, Ju J. Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:5932–7. [PubMed: 15829589]
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001; 409:928–33. [PubMed: 11237013]
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977; 74:5463–7. [PubMed: 271968]
- Shearer AE, Deluca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, Scheetz TE, Smith RJ. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2010
- Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology*. 2008; 26:1135–45.
- Shoubridge C, Tarpey PS, Abidi F, Ramsden SL, Rujirabanjerd S, Murphy JA, Boyle J, Shaw M, Gardner A, Proos A, Puusepp H, Raymond FL, Schwartz CE, Stevenson RE, Turner G, Field M, Walikonis RS, Harvey RJ, Hackett A, Futreal PA, Stratton MR, Gecz J. Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability. *Nature genetics*. 2010; 42:486–8. [PubMed: 20473311]
- Simpson MA, Irving MD, Asilmaz E, Gray MJ, Dafou D, Elmslie FV, Mansour S, Holder SE, Brain CE, Burton BK, Kim KH, Pauli RM, Aftimos S, Stewart H, Kim CA, Holder-Espinasse M, Robertson SP, Drake WM, Trembath RC. Mutations in NOTCH2 cause Hajdu-Cheney syndrome,

- a disorder of severe and progressive bone loss. *Nature genetics*. 2011; 43:303–5. [PubMed: 21378985]
- Smith RJ, Bale JF Jr, White KR. Sensorineural hearing loss in children. *Lancet*. 2005; 365:879–90. [PubMed: 15752533]
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458:719–724. [PubMed: 19360079]
- Summerer D, Wu H, Haase B, Cheng Y, Schracke N, Stahler CF, Chee MS, Stahler PF, Beier M. Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing. *Genome research*. 2009; 19:1616–21. [PubMed: 19638418]
- Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, Daly MJ, Morton CC, Gusella JF. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet*. 2011; 88:469–81. [PubMed: 21473983]
- Tang WS, Qian D, Ahmad S, Mattox D, Todd NW, Harrison H, Huang S, Li Y, Wang Y, Li HW, Lin X. A low-cost exon capture method suitable for large-scale screening of genetic deafness by the massively-parallel sequencing approach. *Genetic Testing & Mol Biomarker*. 2011 in press.
- Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Human molecular genetics*. 2010; 19:R145–51. [PubMed: 20705737]
- Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, Kotsopoulos SK, Samuels ML, Hutchison JB, Larson JW, Topol EJ, Weiner MP, Harismendy O, Olson J, Link DR, Frazer KA. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature biotechnology*. 2009; 27:1025–31.
- Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, Garraway LA, LaFramboise T, Lee JC, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong KK, Borras AM, Gutmann EJ, Dragnev KH, DeBiasi R, Chen TH, Glatt KA, Greulich H, Desany B, Lubeski CK, Brockman W, Alvarez P, Hutchison SK, Leamon JH, Ronan MT, Turenchalk GS, Egholm M, Sellers WR, Rothberg JM, Meyerson M. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat. Med*. 2006; 12:852–855. [PubMed: 16799556]
- Tsurusaki Y, Osaka H, Hamanoue H, Shimbo H, Tsuji M, Doi H, Saito H, Matsumoto N, Miyake N. Rapid detection of a mutation causing X-linked leucoencephalopathy by exome sequencing. *J Med Genet*. 2011; 48:606–9. [PubMed: 21415082]
- Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*. 2009; 6:315–6. [PubMed: 19349981]
- Volpi L, Roversi G, Colombo EA, Leijsten N, Concolino D, Calabria A, Mencarelli MA, Fimiani M, Macciardi F, Pfundt R, Schoenmakers EF, Larizza L. Targeted next-generation sequencing appoints c16orf57 as clericuzio-type poikiloderma with neutropenia gene. *Am J Hum Genet*. 2010; 86:72–6. [PubMed: 20004881]
- Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet*. 2010; 87:90–4. [PubMed: 20602914]
- Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Li RQ, Tang BS. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain*. 2010; 133:3510–8. [PubMed: 21106500]
- Weise A, Timmermann B, Grabherr M, Werber M, Heyn P, Kosyakova N, Liehr T, Neitzel H, Konrat K, Bommer C, Dietrich C, Rajab A, Reinhardt R, Mundlos S, Lindner TH, Hoffmann K. High-throughput sequencing of microdissected chromosomal regions. *Eur J Hum Genet*. 2010; 18:457–62. [PubMed: 19888302]
- Wesolowska A, Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R. Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. *Leukemia*. 2011; 25:1001–1006. [PubMed: 21415851]

- White KR. The current status of EHDI programs in the United States. *Ment Retard Dev Disabil Res Rev.* 2003; 9:79–88. [PubMed: 12784225]
- Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med.* 2011; 13:255–62. [PubMed: 21173700]
- Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature genetics.* 2011; 43:309–U51. [PubMed: 21399634]
- Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 2011; 38:95–109. [PubMed: 21477781]
- Zheng J, Miller KK, Yang T, Hildebrand MS, Shearer AE, DeLuca AP, Scheetz TE, Drummond J, Scherer SE, Legan PK, Goodyear RJ, Richardson GP, Cheatham MA, Smith RJ, Dallos P. Carcinoembryonic antigen-related cell adhesion molecule 16 interacts with alpha-tectorin and is mutated in autosomal dominant hearing loss (DFNA4). *Proceedings of the National Academy of Sciences of the United States of America.* 2011; 108:4218–23. [PubMed: 21368133]

Highlights

1. How to use NGS technologies for studies on epidemiological scales is still an unsolved issue.
2. Gene capture combined with NGS allows much larger number of samples to be examined.
3. Therefore, targeted gene enrichment followed by NGS may provide a solution.
4. Such an approach promises to bring a paradigm shift to studies of genetics of deafness.
5. A review on principles and applications of gene enrichment and NGS technologies is presented.

Table 1

Common bioinformatic databases used for analyzing variants identified by NGS

<i>Analysis questions</i>	<i>Databases available</i>
To find out if the variants have been detected previously or reported in the literature for a particular disease	HGMD (http://www.hgmd.cf.ac.uk/ac/index.php) dbSNP http://www.ncbi.nlm.nih.gov/projects/SNP/
To find out if it is a coding region variant that will result in deleterious effect on protein structure and function	SIFT (http://sift.jcvi.org) DNAnexus (http://DNAnexus.com) GenomeQuest (http://www.genomequest.com) Otoscope (http://www.healthcare.uiowa.edu/labs/morl/) Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim) Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) MAPP (http://mendel.stanford.edu/sidowlab/downloads/MAPP/index.html) Panther (http://pantherdb.org) Mutation Taster (http://www.mutationtaster.org) nsSNP Analyzer (http://snpanalyzer.uthsc.edu) polyPhen (http://genetics.bwh.harvard.edu/pph2/) SAPRED (http://sapred.cbi.pku.edu.cn)
To find out if the variant is located on transcript or CDS region.	RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/)