의학박사 학위논문

Genetic analysis on familial nonsyndromic hearing loss using next-generation sequencing

차세대 염기서열 분석을 이용한 가족성비증후군난청의 유전분석

2014 년 2 월

서울대학교 대학원

의과학과

박 기 범

A thesis of the Degree of Doctor of Philosophy

차세대 염기서열 분석을 이용한 가족성비증후군난청의 유전분석

Genetic analysis on familial nonsyndromic hearing loss using next-generation sequencing

February 2014

The Department of Biomedical Sciences,

Seoul National University

Graduate School

Gibeom Park

Genetic analysis on familial nonsyndromic hearing loss using next-generation sequencing

by
Gibeom Park

A thesis submitted to the Department of Biomedical Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Medical Science at Seoul National University Graduate School

December 2013

Approved by Thesis Committee:

Professor _____Chairman
Professor _____Vice chairman
Professor _____
Professor _____
Professor _____

ABSTRACT

Introduction: Hearing loss is a common sensorineural disorder and half of hearing loss is derived from genetic causes. Due to complexity of hearing mechanism, about 50 genes have been found to have genetic aberrations from previous studies. Next-generation sequencing (NGS) using massively parallel sequencing of DNA is a technology for detection of genetic information and variations from large datasets. I could apply several analytical methods from this platform to pinpoint the causal variations of hearing loss.

Methods: Two independent approaches were used for different familial nonsyndromic hearing loss cases. Whole exome sequencing (WES) for one large family and targeted re-sequencing (TES) for a group of probands of hearing loss were used, respectively. WES was performed with four patients and four normal members in the family and multiphasic analysis was conducted upon copy number variation(CNV), linkage analysis and single nucleotide variation (SNV) calling. TES was performed with 20 probands after pre-screening and I filtered the candidate variants from TES by the multistep criteria to elucidate the possibly causal variants.

Results: In WES study, there were no candidate CNVs that co-segregated with the disease. I detected six regions that had LOD score > 1 from linkage analysis and one SNV in gamma-actin 1 (ACTG1) was a novel mutation. ACTG1 was a known causal gene of hearing loss and the mutation site was predicted to interact with ATP. In TES study, 32 cases were originally collected. Due to failure of pre-screening genetic analysis, 20 cases were

analyzed by TES with SNV and indel in the targeted genes. I filtered the candidate variants by multi-steps with inheritance pattern of the gene, Sanger sequencing of the case, the control and/or the family members and clinical features. 13 of 20 cases had final candidate variations and 25 of total 32 cases had candidate variations in sum.

Conclusions: Trial for potential therapeutics and genetic counseling can be considered, if precise genetic diagnosis for hearing loss is possible. Especially, with analysis of NGS technology, the genetic diagnosis of hearing loss could be more accessible and efficient.

* This work is published in **"**Multiphasic analysis of whole exome sequencing data identifies a novel mutation of ACTG1 in a nonsyndromic hearing loss family. BMC Genomics. 2013 Mar 18;14:191" and "Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss. PLoS One. 2013 Aug 22;8(8):e68692."

Keywords: Hearing loss, next-generation sequencing, whole exome sequencing, targeted re-sequencing

Student number: 2010-21905

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

LIST OF ABBREVIATIONS

NSHL : nonsyndomic hearing loss

NGS : next-generation sequencing

WES : whole exome sequencing

TES: targeted resequencing

WGS : whole genome sequencing

SNV : single nucleotide variation

CNV : copy number variation

AD : autosomal dominant

AR : autosomal recessive

LOD : logarithom of odds

MAF : minor allele frequency

PDB : protein data bank

3D : three-dimentional

nBCNC : narrow bony cochlear nerve canal

GENERAL INTRODUCTION

1. Genetics of hearing loss

Some degree of hearing loss occurs in 1 to 6 per 1000 newborn infants [1]. And hearing loss is the most common sensorineural disorder with increasing prevalence into adolescence [2]. About half of hearing loss is considered as genetic and can be detected by sequencing to help the clinicians and patients manage with it. Hearing loss occurring in early childhood can affect the linguistic development, so it is quite important to improve our techniques to find genetic alterations in patients for further clinical care of this disease.

Hearing mechanism is intricate with the complex auditory systems and with the functions of multiple genes. Hence, genetic hearing loss has heterogeneity with inheritance pattern and causative genes. Classification of loci linked to hearing loss depends on the inheritance mode – DFNA (dominant), DFNB (recessive), DFNX (x-linked), DFNY (y-linked) and DFNM (modifier) [3]. > 100 loci have been mapped for monogenic hearing loss. And the function of known causative genes (~50) are grouped in several categories – 1) regulators of genes (e.g. transcriptional factor), 2) regulator of fluid homeostasis (e.g. transporter and channel), 3) components of junctional complex or tight junctions, 4) extracellular matrix genes, 5) hair bundle organization and function and 6) some neuronal genes [2, 3].

Hearing loss is classified to either syndromic or nonsyndromic. Syndromic hearing loss accompanies other medical problems such as malformations or functional problems except hearing loss. The clinician screens them by multiple tests, such as inspection, physical tests, laboratory tests and imaging. Usher syndrome, Stickler, Waardenburg syndrome, Alport, and Neurofibromatosis type 2 have hearing loss as one of the symptoms of them. Nonsydromic hearing loss (NSHL) consists 70% of inherited hearing loss and most are autosomal recessive up to 80%, in comparison to 20% autosomal dominant and less than 1% X-linked or mitochondrial disorders [4]. Over the decades, many genes were identified to have causative variants by linkage analysis, Sanger sequencing and animal model. But there are still loci that have no known causative genes after linkage analysis. Specifically, GJB2 is the most frequent gene in NSHL.

Hearing loss is often monogenic and genetic test can accurately predict the deafness phenotype. But a candidate gene approach using conventional sequencing related to a certain phenotypic marker can cover only 10~20% of familial NSHL cases. In that prediction of the efficacy on certain therapeutic approaches, e.g. cochlear implantation, is allowed by knowing the cause, more efficient method to find the causative genes is needed.

2. Next-generation sequencing to study NSHL

Automated sequencers based on Sanger sequencing were devised in 1986 and used for Human Genome Project. Another new sequencing technology with massively parallel sequencing that is not Sanger sequencing has been generalized and is called next-generation sequencing (NGS). The procedure of NGS is amplification of DNA by PCR, followed by processing of images from camera. Three major platforms exist – 454, Illumina and SOLiD. 454 and SOLiD use emulsion PCR and Illumina uses solid-phase amplification. Each platform has its own characteristics in detail, but this will not be discussed.

The sequencer reads signal from massive number of DNA fragments (variable length) and this enables the analysis of big data from diverse DNA or RNA origins. Whole genome sequencing, whole exome sequencing, targeted resequencing, pooled sequencing, RNA sequencing, microRNA sequencing, bisulfite sequencing amd ChIP-seq are examples for application of NGS. And most biological samples can be applied to NGS – rare disease, cancer, common disease, population genetics, comparative genetics, etc. Bioinformatic analysis on "big data" is essential to have the results and there are some approaches – *de novo* assembly, alignment on reference, variant detection (SNV, indel, CNV, structural variation and differential expression / methylation / enrichment), annotation and systemic / functional approach.

In study of rare diseases, causal variants of genomic DNA can be found in the global genome or in the particular regions of genome. Hundreds of rare diseases have been elucidated with their causal variants using NGS [5]. Mendelian inheritance in the family should be considered for segregation or

linkage analysis. And the functional relevance of candidate genes should be guaranteed before or after the analysis.

Genetic testing can often predict the deafness phenotype. Genetic testing using NGS could provide an answer to eliminate the need for further expensive testing. Identification of genes and genetic testing will also allow the specific cause of a patient's hearing loss to be uncovered, which cannot be identified by Universal Newborn Hearing Screening. NGS can improve us to find the cause of hearing loss, which allows clinicians to predict the efficacy of certain therapeutic approaches for further therapeutics and protective medications.

3. Whole exome sequencing and targeted resequencing

After target enrichment with exome, the sequencer produces massive data of short reads from sequencing in whole exome sequencing (WES). In Illumina platforms, .fastq files are generated with read names, base sequences and base qualities. Human reference genome is public data as ~3Gb sequences of genomic DNA from one normal human individual. And polymorphisms and variations among other individuals are organized as public database and frequently updated in dbSNP, 1000 genome project, etc. Read alignment procedure of .fastq files on human reference genome is based on sequence similarity between the two sequences. MAQ, CASAVA, BWA, BOWTIE, SOAP and GSNAP are famous alignment programs in NGS. With

information of aligned reads, the depth of coverage in the target regions is checked. To detect single nucleotide variations (SNV) and short indels, the variant calling programs such as Samtools, GATK, VarScan, MuTect, SNVMix, etc. are used with information of mutual locations and the partial differences between reference genome and aligned sequences.

Conventionally, WES is used for detection of SNVs and short indels. But its application can be broaden with searching copy number variations (CNV) from quantitation of read numbers in the target regions and linkage analysis from detected SNVs. In Chapter 1 of this paper, the extended analysis with WES was utilized to elucidate the causal variant of familial NSHL.

In case that we know the specific gene group to analyze, targeted resequencing method can be used for cost-efficiency. Targeted gene capture the following advantages – 1) significant cost saving, 2) higher sequencing accuracy because of deeper achievable coverage, 3) a significantly shorter turnaround time, 4) a more feasible data set for a bioinformatic analysis outcome that is functionally interpretable [6]. Custom-designed probes for target enrichment are in need and the procedures of experiment and analysis are similar to WES. In Chapter 2 of this paper, 20 familial NSHL cases were analyzed with targeted resequencing.

4. Technical purposes of this study

Clinical studies for monogenic disorders often use WES and this has been successful for already published studies whose number is >100. It is

necessary to carefully look at the present typical utilization of WES for monogenic disorders.

In WES studies, some limitations exist compared to whole genome sequencing (WGS). First, though we can focus on the exon regions with deeper coverage than WGS, incompleteness of capturing experiments still exists. This makes some regions in WES rather disadvantageous for analysis with lower or no coverage. These missing regions make us unknown whether causative variations are located in the missing regions or the phenotype is just not derived from SNV or indel, in case of the failed situations of WES analysis. This situation can be the reason or just an excuse for failed cases, but it is often hard to determine which of them is correct. Second, the probable regions for causative variations, such as promoters, non-coding RNAs, enhancers, etc. are omitted in WES. Because not many study groups have WGS data for clinical samples, we cannot know how frequent the causative variants are located out of the exons. This makes us again hard to try WGS for clinical samples, due to tough consideration between the scale of region size and inefficient coverage, running time and cost. It is sometimes adventurous to utilize WGS, but variations out of exons could exist, supposedly not found by WES.

However, we still stick to WES in many cases, which is somewhat inevitable. We have a lot of clinical samples for NGS analysis and it is still reasonable to think that most genetic causes of phenotype are inside exon regions. And not captured regions compose only several percentages of the whole exon regions. Additionally, >90% of reads from non-coding regions of

WGS data might be waste, in case of coding cause of the phenotype. There are still many failed cases of WES, but we hope to succeed in the NGS analysis for these desirable reasons of WES. And WES has already produced many published outcomes.

Previous studies have focused on only SNVs and indels in WES analysis. But sequence data and quantitative data are both outcomes of WES and there are additional potentials for using the both to take other approaches except SNV calling – e.g. copy number and linkage analysis. I extended the conventional method of WES analysis, that is, SNV and indel calling to more informative analysis – multiphasic analysis on copy number variation, linkage analysis and final cosegregation study. The major points of the study in Chapter 1 are 1) possible methods for WES using sequence and quantity were used in sum, 2) the outcomes were reliable with correlation to known CNV genes and linkage by SNPchip, 3) thus, these methods could substitute for the previous array-based genomic data, 4) up-to-date methodology in WES analysis on CNV was applied and 5) the final causative mutation was found. Detailed aspects of this new multiphasic approach will be discussed in Chapter 1.

On the other hand, if we know the putative causative genes, targeted resequencing is beneficial for cost and read depth. Disorders caused by heterogenous genes, such as hearing loss, retinitis pigmentosa, bone malformation, and even cancer are good examples of conditions for targeted resequencing. Instead of sequencing ~20,000 genes, we can sequence just tens to hundreds genes with high coverage. Or, when one or two genes are too long

for Sanger sequencing, especially when dealing in multiple samples, we can efficiently sequence the whole region of the genes with low cost. Due to several benefits of targeted resequencing, this method is suitable for quick clinical applications of NGS technology, in case the characteristics of disorder are proper to it. But the existence of possible false positives and negatives should be solved for clinical diagnosis.

False positives and negatives in NGS analysis are not ignorable in the basic researches and are major obstacles in clinical trials. They come from several issues – sample qualities, DNA or RNA degradation, target enrichment, PCR duplicates, low base qualities, GC content, sequence error, repetitive sequence of reference, existence of paralogs, etc. In RNA-seq, the complexity increases due to splicing events of mRNAs. For routine clinical use, NGS may be combined to the conventional methods, such as Sanger or TaqMan genotyping. And some analysis on the efficacy of targeted resequencing was conducted in this study – capture efficiency, correlations of coverage, SNV number between the groups with or without the final candidates. Because targeted resequencing is newer method in NGS study, its performance is not well-known. Thus, my study on targeted resequencing shows the reliable performance of it upon clinical samples. More trials need to be done in the future, and my study can be the good start in targeted resequencing, especially upon genetic hearing loss.

5. Biological and clinical implications of this study

Though NGS studies tend to be global genomic search and observation with less conventional hypothesis and experiment paradigms, ground hypothesis should be well-designed before the production of data. Simple profiling of samples is often acceptable, but sometimes makes the researchers hard to interpret the results. What to analyze comes from what to look at and what to look at comes from biological or clinical implications related to the experiments. For example, when analyzing cancer genomics, we can increase the number of samples for searching common variations or for subgrouping the samples, both of which are good for elucidation of the cancer biology and application to therapeutic field. Another example is proper selection of patients groups and normal groups. Some interesting diseases without well-known molecular mechanisms are good to analyze and some alternative approaches on well-known diseases are also plausible. In this sense, conversation between researchers and clinicians is also important. Clinicians who do not understand the research or vice versa are both not desirable. The third example is targeted resequencing. The root hypothesis is that recurrent mutations which have been observed from many studies are expected to reemerge in our study. And genes that have relevant biological meanings can be included, which leads researchers to think about biological implications of the experiments and genes.

The clinical purpose of this study is that it is more effective and efficient to find the genetic cause of hearing loss when we combine several tools including various parts of NGS and/or some conventional tools. I combinded copy numer detection, linkage analysis and cosegregation study in

the single analysis of WES data in Chapter 1. Also, I made up the clinical outlines for genetic hearing loss, with combination of Sanger sequencing by phenotypes and targeted resequencing in Chapter 2. Early and accurate diagnosis is important for linguistic development and proper therapeutic approaches. Easier and more accurate diagnosis can be done with these new tools using NGS technology.

Furthermore, some biological implications exist in this study. 3D structures of protein greatly influence its function and binding preferences to other molecules. And domains composed of specific sequence patterns or some residues that are found in common among the protein families, imply the functional role or binding partners of the protein. ACTG1 (gamma-actin 1) in Chapter 1, uses ATP as an energy source for polymerization to form fibers. I found the data in PDB (Protein Data Bank) database that the causative mutation of ACTG1 binds to ATP directly, which was expected by software. And LCCL domain of COCH (cochlin) whose impairment is substantial for DFNA9, had the candidate mutation in targeted resequencing. And some additional genes will be discussed in Chapter 2.

6. Previous studies on genetic hearing loss using NGS

There were at least three articles on genetic hearing loss using NGS, when I searched Pubmed website (http://www.ncbi.nlm.nih.gov/pubmed). The first one was about discovery of the candidate mutation at SMPX gene using

NGS analysis after enrichment of X chromosome, in the family of X-linked inheritance [7]. They performed linkage analysis with microsatellite markers before the NGS study and did Sanger sequencing of three genes near LOD peaks. But no candidate mutations were found. So, they did NGS analysis and found the mutation of SMPX. The points different from my study were the form of Mendelian inheritance that could decrease the target regions, the prior linkage analysis by microsatellite markers that had failed to find the mutation and the expression analysis that they conducted to validate the expression in the inner ear, because SMPX was a novel causative gene. My study emphasizes that WES analysis alone with multiphasic approach can be a tool enough to find the novel mutation in rare diseases, in comparison to their previous study.

The second and the third ones were about targeted resequencing. One group used 15 genes related to hearing loss and performed analysis on 5 patients [8]. Another group used 80 genes on 8 families with hearing loss [9]. The points different from my study were that the number of patients was much greater in my study rather than in their studies and that I combined targeted resequencing with the conventional diagnostic method in sum, to make the methodology more available to the clinical applications. And with the greater number of patients, the detection rate did not decrease abruptly – 65% (13/20) in my study, and 4/5 or 7/8 in the previous studies.

CHAPTER 1

Multiphasic analysis of whole exome sequencing data identifies a novel mutation of ACTG1 in a nonsyndromic hearing loss family

INTRODUCTION

The recent development of NGS technologies with massively parallel DNA sequencing presented the easier and more accurate access to genetic and genomic composition of the human genome to the researchers than ever. As many categories as possible are being tried for applications of NGS to biological field. These applications include the extended versions of classical genetics and the novel methods that could not be fulfilled by the conventional approaches. But, the novel aspects must exist in any NGS analysis, and the tools for analysis have been being developed steadily. If we focus on the coding regions for the analysis and we do not know which genes are the targets, we need to profile the whole exome sequences, which is now available owing to NGS. Using the advantage of exome sequencing, causal variants of numerous disorders have been identified. Exome sequencing can be used to identify disease-causing single nucleotide variations (SNVs) in Mendelian disorders with its powerful utility. [1-4]. Despite the successful applications for discovery of many causal genes in Mendelian disorders, analysis of WES data remains challenging [1]. An individual exome has more than 20,000 variants compared with the reference genome. Even in familial Mendelian disorders, the overall success rate for identifying disease-causing genes is around 50% [5]. Thus, it is necessary to analyze the unrevealed parts of Mendelian disorders with additional analytic methods of WES to fulfill the promise of WES for routine diagnosis of Mendelian disorders.

Filtering patient data against normal populations and inferring

identity-by-descent (IBD) in family studies can enrich candidate genes [4,6]. Genetic linkage analysis has also been a powerful tool for isolating potential causal candidate variants. A two-step approach of linkage analysis using single nucleotide polymorphism (SNP) microarrays to detect high logarithm of odds (LOD) score regions and subsequent targeted re-sequencing of regions of interest has been utilized in many genomic studies to intensify the power of detection [7]. Classically, microsatellite markers have been used for linkage analysis, and now millions of SNP markers can be used to provide higher resolution in order to pinpoint candidate loci [8]. Currently, there are many efforts to use coding SNP information from WES data to facilitate genetic linkage mapping. Specifically, coding SNP data from WES can be used to establish multiphasic exome analyses based on linkages and SNVs [4,9].

Copy number variation (CNV) has been implicated in both Mendelian diseases [10] and common diseases such as obesity [11] and schizophrenia [12]. The presence of large insertions or deletions in patients is typically investigated prior to SNV analysis by karyotyping, fluorescence in situ hybridization (FISH), and/or array comparative genome hybridization (aCGH). Estimation of CNV is a challenging aspect of WES analysis, in which local depths of coverage must be mapped to copy numbers. Indeed, aCGH has limitations in detecting high CNV regions. Conversely, CNV data based on WES provides more accurate copy numbers because the depths of exon coverage from WES data vary linearly with real copy numbers [13]. Bioinformatics tools to analyze copy numbers from WES data are now

publicly available [14].

Nonsyndromic hearing loss (NSHL) contributes to more than 70% of inherited cases of hearing loss. To date, approximately 50 genes have been shown to be causally related to NSHL. Many studies have identified more than 129 loci responsible for NSHL; however, 47 loci have not yet been mapped to proper genes [15,16]. The complexity of the auditory system may explain why so many genes and loci are linked to hearing loss. The genetic causes of hearing loss can be detected by sequence analysis, which helps clinicians and patients to delineate the basis of the disease. Given that hearing loss in early childhood can affect linguistic development [16], it is important to improve current techniques for identifying genetic alterations that cause NSHL. Earlier identification of such alterations in patients and families may allow for better clinical management of NSHL.

Analysis of WES data can be expanded to obtain more information useful for identifying causative mutations in Mendelian diseases. In this paper, in order to analyze WES data from an entire family, I applied three different methods, namely, CNV, linkage, and segregation analysis. By combining the results obtained from these methods, I efficiently identified a causative mutation from the family data. Using this multiphasic approach of WES data to a NSHL family, a novel mutation of ACTG1, a known causative gene of DFNA20/26 was identified (Figure 1-1).

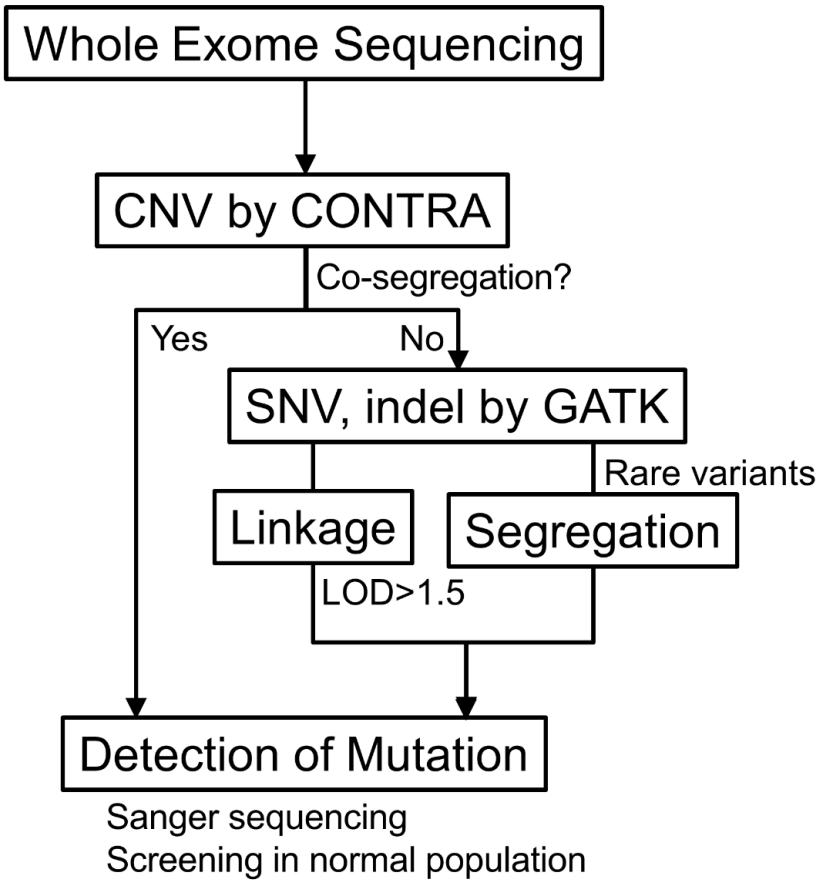


Figure 1-1. Scheme of the multiphasic analysis of WES data

MATERIALS AND METHODS

Subjects

This study was approved by the Institutional Review Boards (IRBs) of Seoul National University Hospital (SNUH) and Seoul National University Bundang Hospital (SNUBH). Written informed consent for participation in the study was obtained from participants or from a parent or guardian in the case of child participants. A three generation pedigree was established for the family (SNUH3) (Figure 1-2). Among the 15 subjects in the SNUH3 family, 13 were willing to participate in this genetic study, while two reportedly deaf subjects (II-8 and III-3) refused participation. DNA from blood lymphocytes was isolated from the 12 subjects, while DNA from III-2 was obtained with a buccal swab.

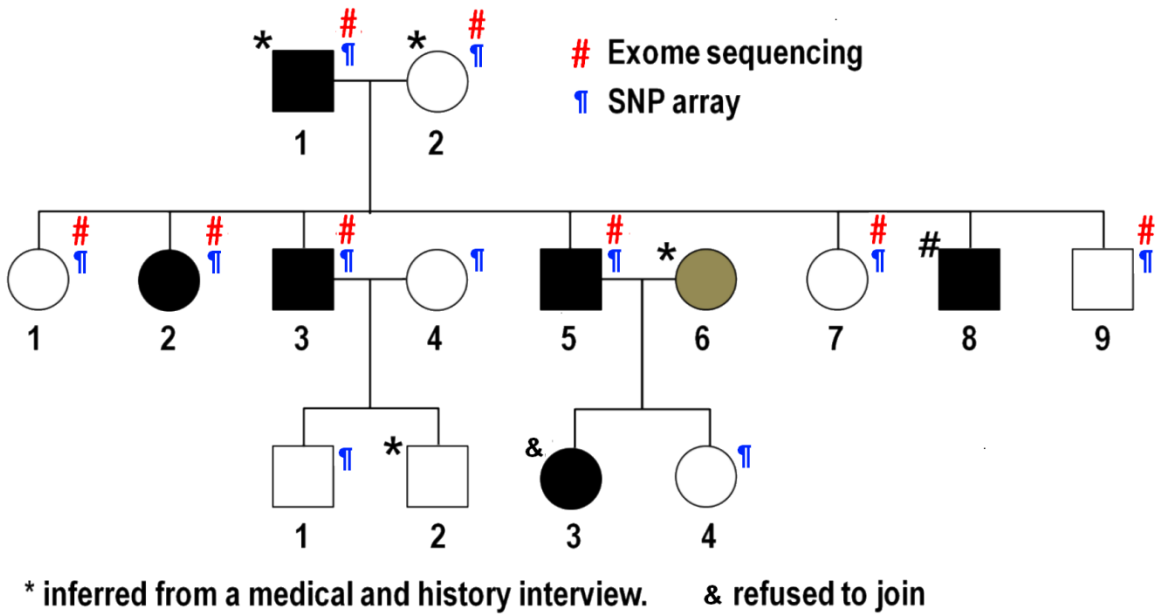


Figure 1-2. Pedigree with phenotype and experimental information. Black and white denote affected and unaffected subjects, respectively, while shaded grey represents equivocal hearing status.

Audiometric analysis

Pure tone and speech audiometry and physical examinations were performed for nine members of the cohort (Figure 1-3). Pure tone audiometry (PTA) with air and bone conduction at frequencies ranging from 250 to 8,000 Hz was obtained from the recruited subjects according to standard protocols. The auditory phenotype was inferred from thorough medical and developmental history interviews from one deaf subject (I-1), two likely unaffected subjects (I-2 and III-2), and one subject (II-6) with an equivocal hearing status.

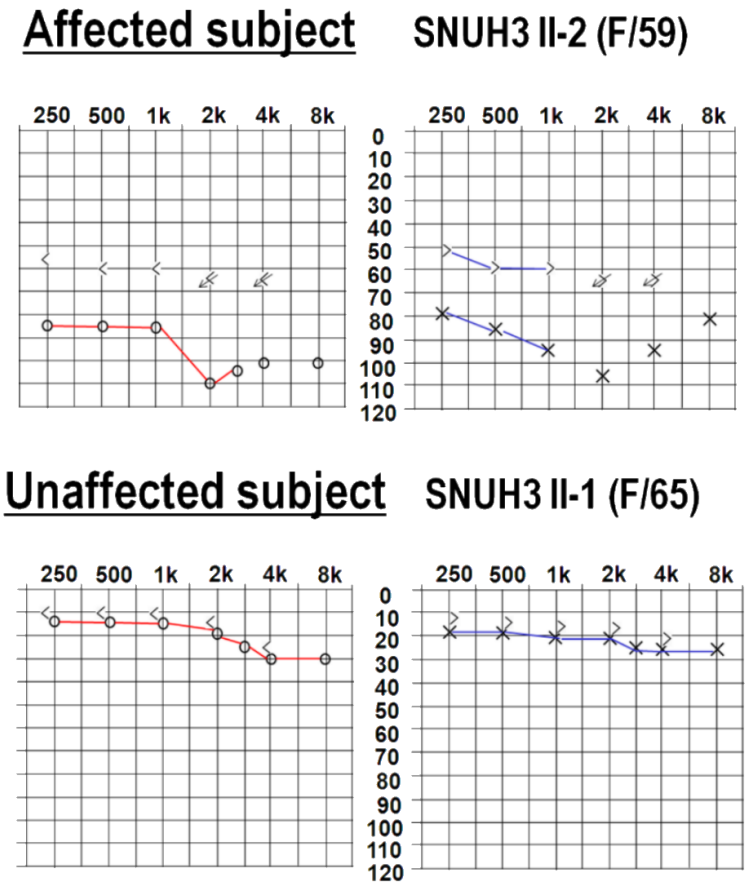


Figure 1-3. Typical audiograms of affected and unaffected subjects

WES

Eight of the 13 recruited subjects (four affected and four unaffected) were chosen for commercial WES (Otogenetics, Norcross, GA) and analyzed as previously reported [3]. Briefly, paired-end reads of 100 bp from the eight subjects were aligned by bwa-0.6.1 to the UCSC hg19 reference genome using default settings. SAMtools and Picard were used to process SAM/BAM files and mark duplicates. Local realignment around indels and base quality score recalibration was done for each sample, and variants were called by a unified genotyper in GATK-1.3. Perl scripts and ANNOVAR were used to annotate variants and search the relevant known SNPs and indels from dbSNP135 and the 1000 Genome database. Variants with a read depth greater than 10 and genotype quality score greater than 30 were filtered for further analysis.

CNV analysis using WES data

CNVs were detected by CONTRA software [36] using BEDTools to calculate coverage per exon and apply statistics to normalize coverage data and test fold changes. A new baseline file was produced using our data, but I expected to detect distinct deletions or amplifications. Polymerase chain reaction (PCR) duplicates were removed by Picard before using CONTRA.

I tabulated a 3 × 2 exon copy variation contingency table based on the whole per-exon CNV status of the eight subjects (Table 1-1). Fisher's

exact test was used to assess the significance of differences between proportions of abnormal copy number events present in affected and unaffected family members. I assumed that all of the subjects were independent in order to conduct an alternative practical method to find loci that segregated with the disease.

Table 1-1. **Exon CNV contingency table**. Based on the whole per-exon CNV status.

Copy number	**Patient**	**Normal**	**Total**
gain	n_{21}	n_{20}	n_{2+}
normal	n_{11}	n_{10}	n_{1+}
Loss	n_{01}	n_{00}	n_{0+}
	n_{+1}	n_{+0}	8

Linkage analysis using WES and SNP microarray data

The algorithms of linkage analysis include the two major methods, Elston-Stewart method and Lander-Green method. I performed parametric linkage analysis with the R package paramlink which uses Elston-Stewart method [37]. The inputs of Elston-Stewart method are a simple pedigree that has no cycles and a single pair of founders, and phenotype information. Using WES data, I filtered out the following variants: those located on sex chromosomes, those with low coverage (<10X), and those with a low genotype quality score (<30) in any of the eight subjects with 17,498 SNVs. I used a Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA), which contains 328,125 SNP markers located on autosomal chromosomes. The pedigree suggested an autosomal dominant mode of inheritance, and thus I assumed an autosomal dominant model with default values of full penetrance (f0, f1, f2) = (0, 1, 1) and disease allele frequency = 1e-05. The penetrance parameters f0, f1, and f2 were also defined using conventional notation as below.

$$f_i = P(\text{affected} \mid i \text{ copies of the disease allele})$$

The recombination fraction between the disease locus and markers was set to θ=0 by default. I computed single-point LOD scores for all markers. I compared LOD scores from SNP microarray and WES. I matched the subjects and the markers that were common between both platforms using manual python and R scripts. Finally, single-point analyses were performed with all of the data.

3D structure of actin gamma-1 (ACTG1)

Protein damage prediction analysis was performed using HumDiv and HumVar in Polyphen2 [17], and also by MutationTaster [18]. The mutation site was visualized using the 3D structure of bovine beta-actin bound by adenosine triphosphate (ATP) with profilin. Bovine beta-actin actin has a 99% identity with human gamma-actin. The ATP binding site was analyzed using the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (www.pdb.org) [38]. PDB entry 2BTF [39] on P60712 (ACTB_BOVIN) with P02584 (PROF1_BOVIN) was downloaded, visualized, and modified by Bioclipse [40] to observe the 305Met residue in 3D. Domain and residue information was searched in PDB.

RESULTS

Clinical features of a NSHL family

A Korean family with six members affected by NSHL and seven unaffected members were identified (Figure 1-1). Pure tone audiometry (PTA) was performed on nine family members, three of whom (II-2, II-3, and II-5) exhibited profound post-lingual hearing loss. The three members had normal cognitive function and no anomalous-looking features. They had no abnormalities in a battery of clinical tests ranging from general physical examinations, chest X-rays, and simple blood tests to detailed imaging studies including brain MRI and temporal bone CT scans, excluding the possibility that hearing loss in these patients was syndromic. The other siblings (II-1, II-7, and II-9) exhibited normal hearing. Patients II-2, II-3, and II-5 estimated that their hearing loss became severe in their 30's, during which time they started to wear hearing aids. Their hearing loss was further aggravated and became profound in their late forties. Ultimately, patients II-3 and II-5 no longer benefited from hearing aids and underwent cochlear implantation. They achieved recognition of common sentences without lip reading one year after implantation. GJB2, one of the most frequently detected genes in individuals with NSHL was first investigated with the sequence in the NSHL patients. After failing to identify any mutations in GJB2, WES on several members of the Korean family was performed in order to identify a disease-causing

mutation.

Copy number analysis using WES data

I performed WES from the parents and six siblings (four affected and four unaffected members, Figure 1-1). The mean coverage of each sample ranged from 40.3X to 51.3X, and 87.0% to 90.5% of the targeted exome had at least 10 reads. A multiphasic WES analysis was designed to find a causative NSHL mutation (Figure 1-3). First, I investigated co-segregation of copy number duplication or deletion in exomes of patients using CONTRA software. I detected five CNV loci with distinct features in the plots (Figure 1-4). None of the CNVs co-segregated with affected or unaffected family members. One CNV locus of the CNVs from three members (high copy number exons in II-9, and low copy number exons in II-3 and II-7) was located in 8p23.1, a region that contains beta-defensin genes and SPAG11 (Figure 1-5). The following genes were identified as being located at regions of distinct CNVs in the indicated family members: GSTM1 in 1p13.3 (I-1, II-3, II-7, and II-9) (Figure 1-5), UGT2B17 in 4q13.2 (I-2 and II-7), BNTL3 in 5q35.3 (II-1), and LILRB2 in 19q13.4 (I-2) (data not shown). I also applied Fisher's exact test for the LOD score per exon to detect co-segregated regions of CNVs, but there were no peaks with values reaching significance. I identified two groups based on the pattern of segregation of SPAG11, GSTM1, and beta-defensin genes to validate the relevance of this method (Figure 1-5).

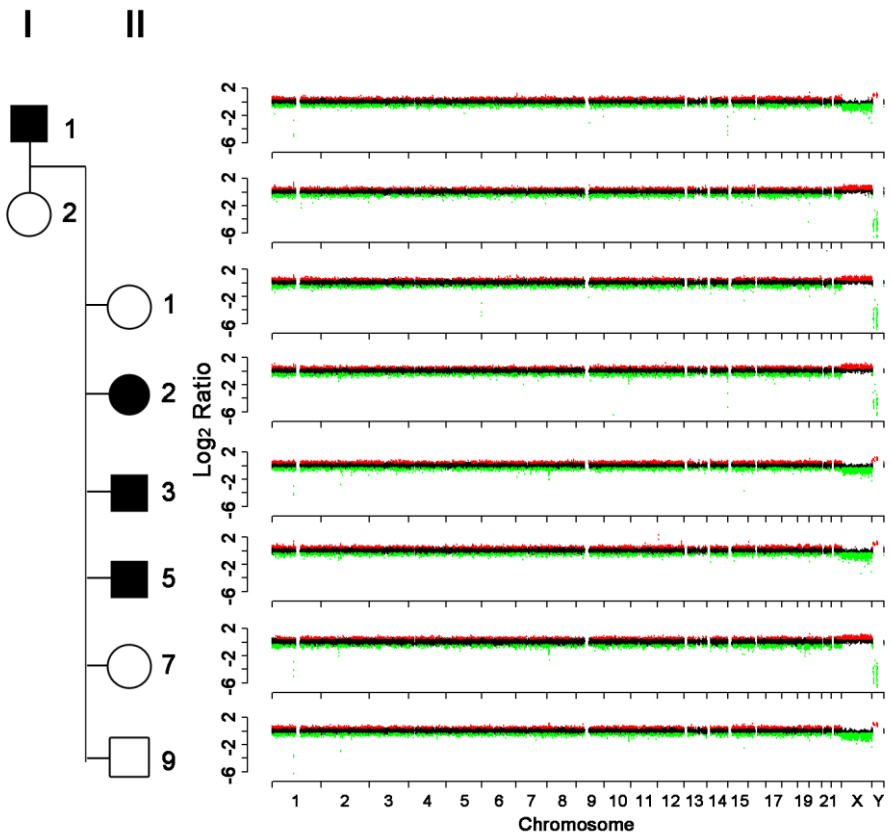


Figure 1-4. **CNV throughout the chromosomes** − 1p13.3, 4q13.2, 5q35.3, 8p23.1, and 19q13.4 have distinct CNVs (14q32.3 is distinct, but contains variable regions associated with antibody production). Red and green dots are exons with $p<0.05$.

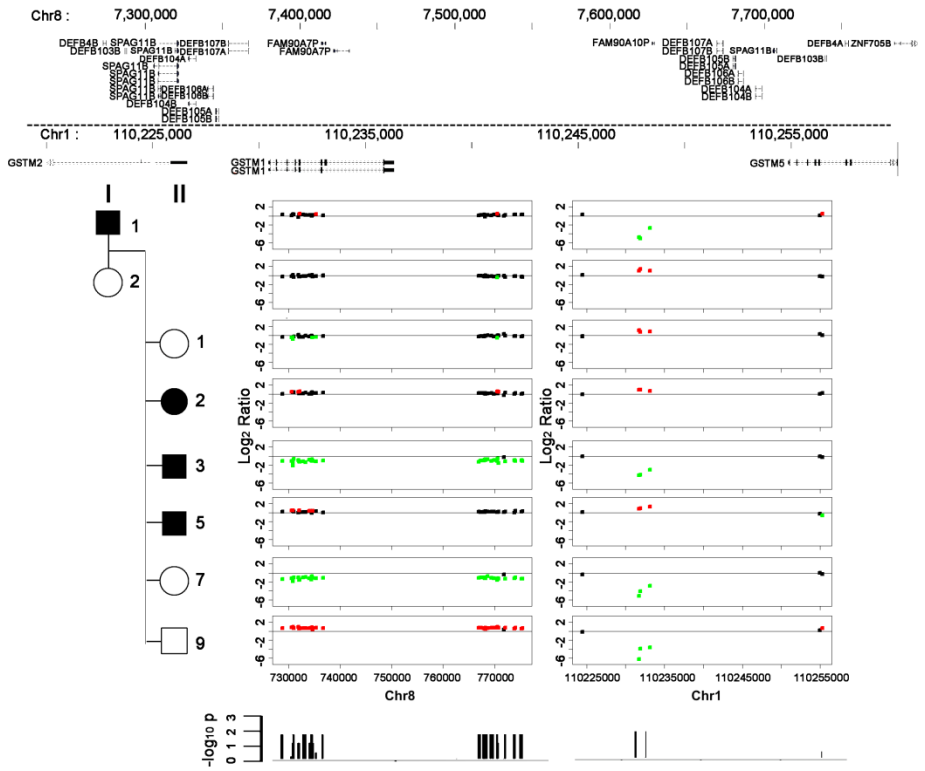


Figure 1-5. **CNV regions detected in the previously known regions**. 8p23.1 containing beta-defensin clusters and 1p13.1 containing GSTM1 of eight subjects. Red and green dots are exons with $p<0.05$ of CNV. Co-segregated regions of CNVs were also analyzed by Fisher's exact test.

Exome linkage analysis

Because the pedigree strongly suggested an autosomal dominant mode of inheritance, I identified 17,498 coding autosomal SNVs from WES data and performed single-point linkage analysis. I identified six hot spots where a number of peaks were closely clustered (Figure 1-6). Specifically, I identified peaks on chromosomes 3, 11, 13, 14, 16, and 17 consisting of 11, 67, 2, 13, 17, and 13 exons, respectively.

I validated single-point linkages using a SNP microarray containing 328,125 SNPs. Along with the eight initial family members recruited for WES analysis, I included three additional subjects (II-4, III-1, and III-4) to validate the significance of peaks obtained from exome linkage analysis. The six hot spots detected from sequencing data were also detected in microarray analysis with a relatively high LOD score (Figure 1-6). Adding three more subjects to the linkage analysis enhanced the peaks at chromosomes 11 and 17, which consisted of one and three SNPs (LOD score >2), respectively. The genotype patterns of these four peaks were perfectly matched with an autosomal dominant mode of inheritance.

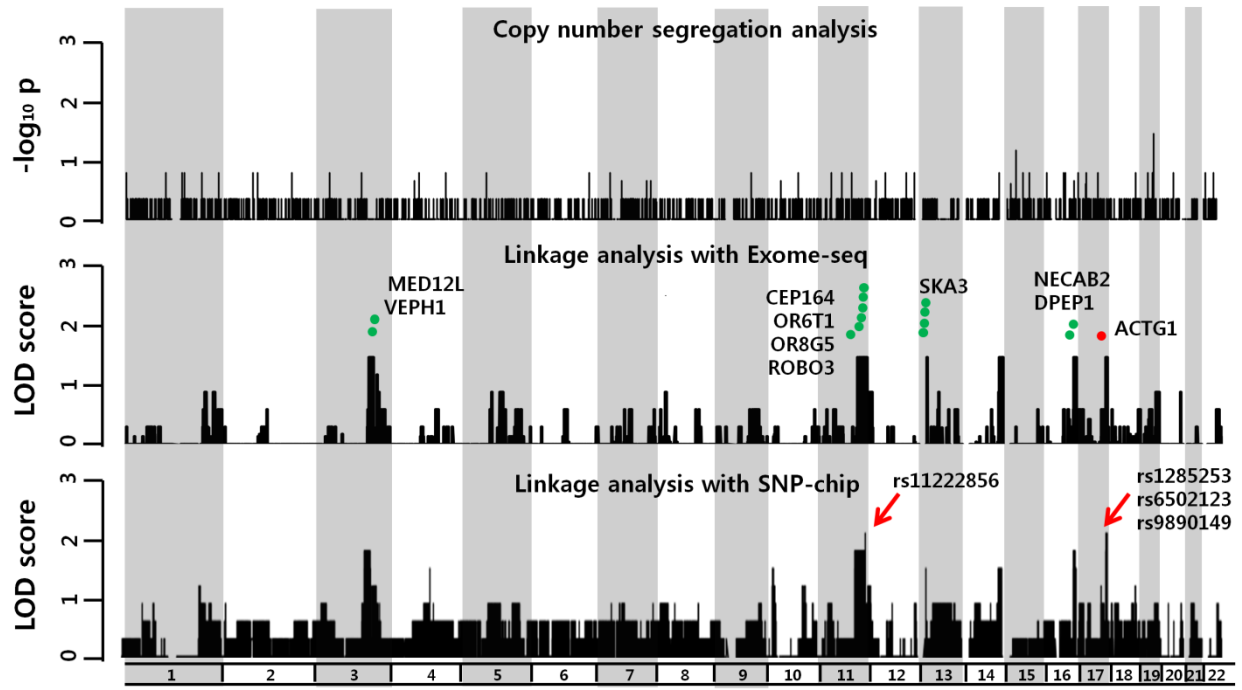


Figure 1-6. Summary of the results from multiphasic analysis. Fisher exact test on CNVs (top). Linkage analysis with SNVs by WES identified six high LOD regions on chr3,11, 13, 14, 16 and 17 (middle). Segregation analysis independently identified 15 SNVs co-segregating with NSHL (green dots) including a novel variant in ACTG1 on chr17 (red dot). Linkage analysis was also performed by SNP microarray with three more subjects in the family as to enhance the true peaks (red arrow) (bottom).

SNV analysis

Based on the WES analysis of four affected and four unaffected family members, I identified 18,748~20,025 SNVs and 413~457 indels. These were reduced to 962~1,123 SNVs and 140~153 indels after filtering through the dbSNP135 and 1000 Genome databases. Fifteen variants causing amino acid changes were selected based on their co-segregation pattern within the family (Table 1-1). All of the 15 variants on chromosomes 3, 11, 13, 16, and 17 corresponded to regions with high LOD scores (Figure 1-6). One novel mutation in actin gamma (ACTG1) was identified, consisting of a methionine to threonine substitution at amino acid 305 (p.M305T) (Figure 1-7). This candidate variant was validated by Sanger sequencing and co-segregated with hearing loss in all family members (Figure 1-8). Many mutations of ACTG1 were identified as causative on genetic hearing loss through several previous studies (Figure 1-7). But p.M305T is a novel mutation, not found in any other researches.

Table 1-2. Cosegregating SNVs and indels. Nonsynonymous SNVs and indels identified in patients but not in non-symptomatic family members

Gene	Chr	Nucleotide variation	Amino acid variation	Frequency in 1,000 genome	dbSNP135
MED12L	chr3	c.G3629A	p.R1210Q	0.23	rs3732765
VEPH1	chr3	c.T1564C	p.S522P	0.28	rs11918974
CWF19L2	chr11	c.A2681G	p.Y894C	0.27	rs3758911
CEP164	chr11	c.G281A	p.S94N	0.19	rs490262
OR6T1	chr11	c.G465C	p.W155C	0.0046	rs150534954
OR8G5	chr11	c.G287A	p.C96Y	0.45	rs2512168
OR8G5	chr11	c.G716A	p.G239E	0.5	rs2512167
ROBO3	chr11	c.G1247A	p.R416H	0.14	rs3862618
SKA3	chr13	c.A1157G	p.K386R	0.13	rs11147976
SKA3	chr13	c.C1142T	p.T381I	0.11	rs11147977
SKA3	chr13	c.G559A	p.V187I	0.14	rs61950353
SKA3	chr13	c.208delC	p.Q70fs	-	rs151272242
NECAB2	chr16	c.C704G	p.T235S	0.2	rs2292324
DPEP1	chr16	c.G1051C	p.E351Q	0.24	rs1126464
ACTG1	chr17	c.T914C	p.M305T	-	-

Figure 1-7. **p.M305T mutation in ACTG1**. The p.M305T mutation reported in this study as well as several other previously reported mutations in ACTG1 cause hearing loss.

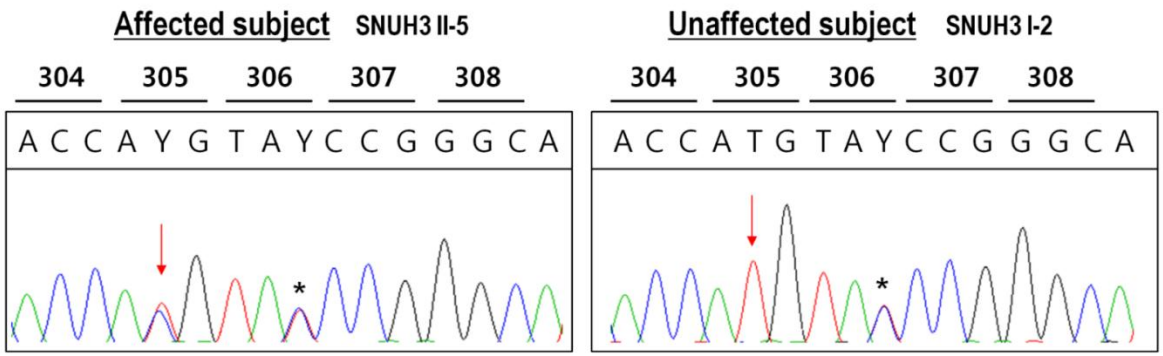


Figure 1-8. **Sanger sequencing of p.M305T**. p.M305T (arrow), confirmed by Sanger sequencing, co-segregated perfectly with hearing loss (asterisk: synonymous SNV)

ACTG1 (DFNA20/26; MIM: 604717) was strictly conserved in 19 of 20 eukaryotes analyzed (HomoloGene:74402), with the M305 codon being conserved in 19 species. Protein damage prediction analysis identified p.M305T as “possibly damaging” by HumDiv, “probably damaging” by HumVar in Polyphen2 [17], and “disease causing” by MutationTaster [18]. The mutation site, Met305, was visualized using the 3D structure of bovine beta-actin bound by adenosine triphosphate (ATP) with profilin (Figure 1-9). There was no 3D structure of the whole human ACTG1 that was analyzed by crystallography. But there are several homologs of human ACTG1, because actin family has >90% similarities between each other and inter-species conservation of actins are strongly conservative. 3D structures of alpha- and beta-actins from several species have been characterized, published and uploaded in database, such as PDB (protein data bank). I selected bovine beta-actin, because it is a cytoplasmic, “non-muscle” actin like human ACTG1 and the protein sequence similarity between them was 99% (371/375) by protein blast alignment (Figure 1-10). In many times, 3D structures of protein by crystallography come with protein complex, and beta-actin was bound to profillin for the characterization of 3D structure. And one ATP molecule was also bound to beta actin. There was one functional domain, “Actin-like ATPase domain” (Figure 1-11). I checked Met305 was located in the binding sites of either profillin or ATP. The methionine was closely located to the ATP molecule. In addition, Met305 is listed as a predicted residue for the ATP binding site by the Protein Data Bank (PDB) (Figure 1-12).

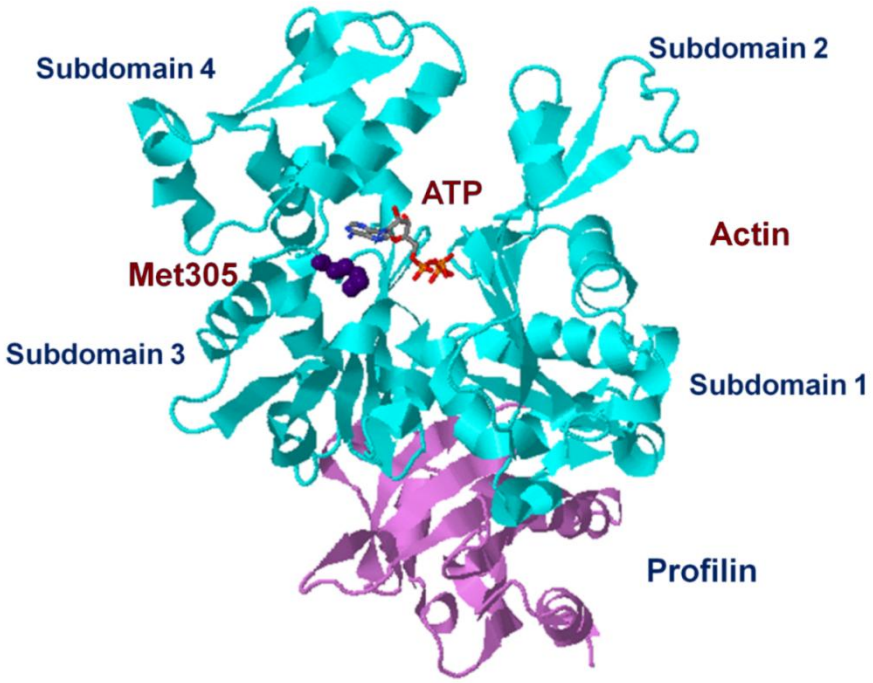
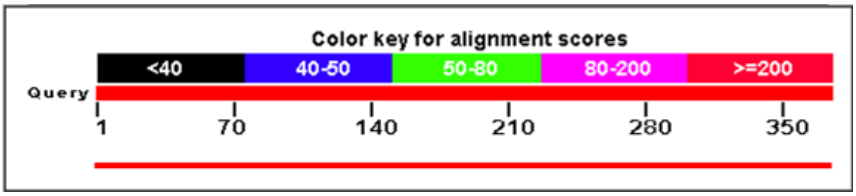


Figure 1-9. **3D structure of bovine beta-actin**. Bovine beta-actin has >99% similarity in protein sequence of ACTG1. Met305 is located close to the ATP binding site of bovine beta actin.




```
Query= sp|P60712|ACTB_BOVIN Actin, cytoplasmic 1 OS=Bos taurus GN=ACTB PE=1
SV=1

Length=375
                                                                      Score     E
Sequences producing significant alignments:                           (Bits)  Value

lcl|43395  sp|P63261|ACTG_HUMAN Actin, cytoplasmic 2 OS=Homo s...     780    0.0

ALIGNMENTS
>lcl|43395 sp|P63261|ACTG_HUMAN Actin, cytoplasmic 2 OS=Homo sapiens GN=ACTG1
PE=1 SV=1
Length=375

 Score =   780 bits (2015),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 371/375 (99%), Positives = 375/375 (100%), Gaps = 0/375 (0%)

Query  1    MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS  60
            M+++IAALV+DNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS
Sbjct  1    MEEEIAALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS  60

Query  61   KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  120
            KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT
Sbjct  61   KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  120

Query  121  QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL  180
            QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL
Sbjct  121  QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL  180

Query  181  AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY  240
            AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY
Sbjct  181  AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY  240

Query  241  ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTVLS  300
            ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTVLS
Sbjct  241  ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTVLS  300

Query  301  GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ  360
            GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ
Sbjct  301  GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ  360

Query  361  EYDESGPSIVHRKCF  375
            EYDESGPSIVHRKCF
Sbjct  361  EYDESGPSIVHRKCF  375
```

Figure 1-10. **Sequence similarity of ACTG1 and bovine beta-actin.** Protein sequence homology was calculated by protein blast.

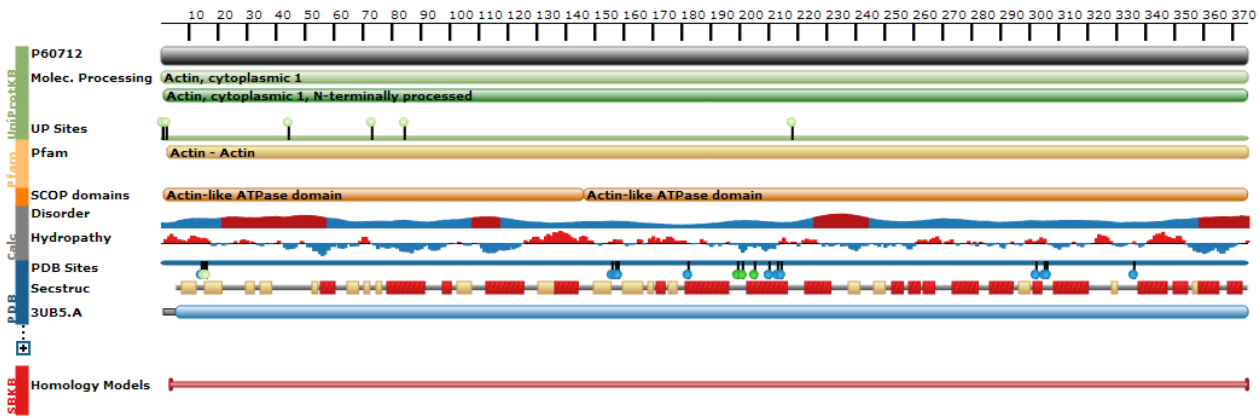


Figure 1-11. **Domains and residues of bovine beta-actin**. They are depicted in Figure 1-12 in more detail.

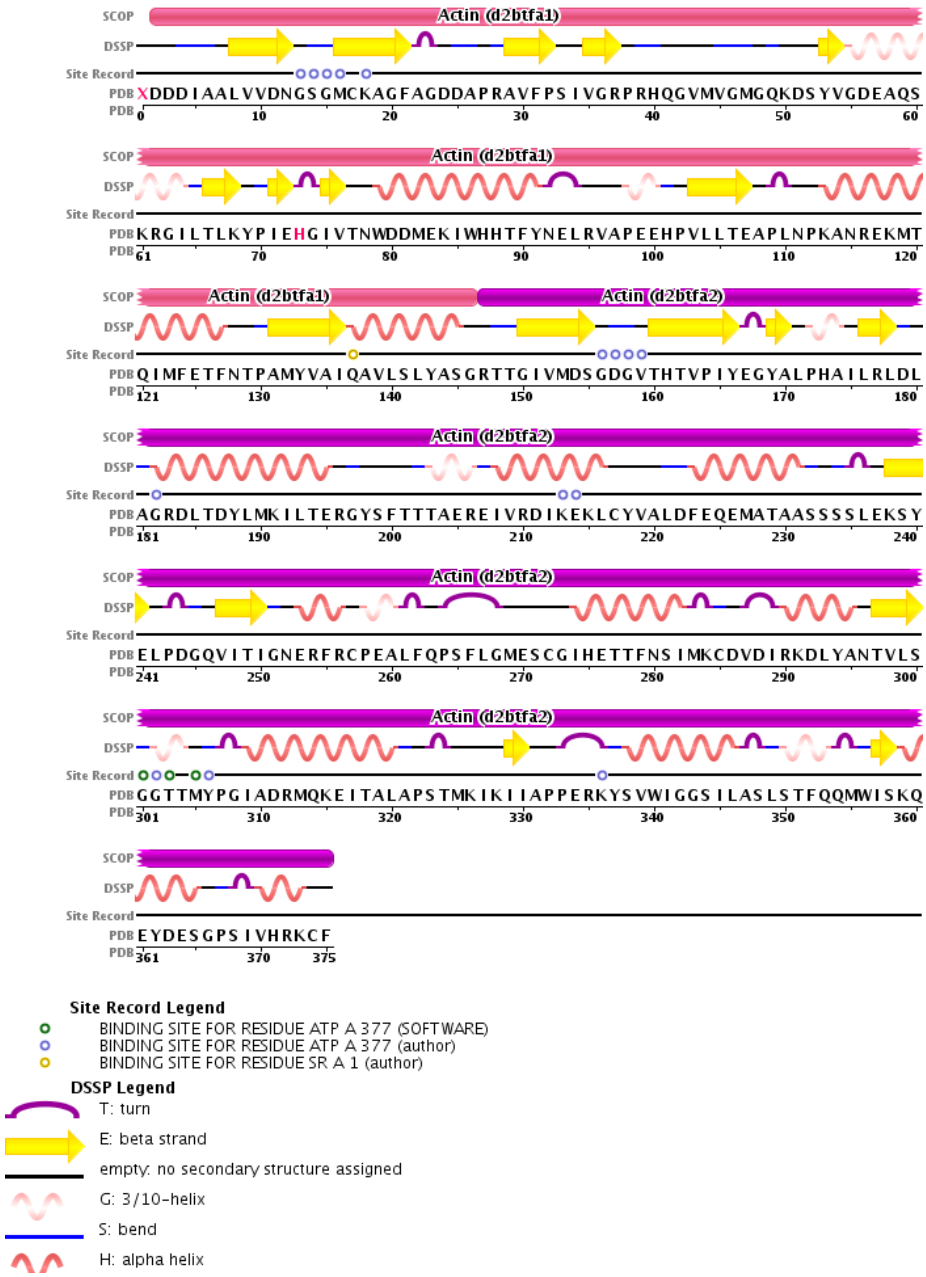


Figure 1-12. **Linear structure of bovine beta-actin**. Notice that Met305 is a predicted site for binding to ATP by software.

DISCUSSION

WES has been used for identifying the novel causative mutations of Mendelian disorders in more than 180 genes [41]. But it is still necessary to increase efficiency of gene discovery using WES data. In this regard, I analyzed WES data from a family with a history of NSHL by focusing on three categories of genetic information: CNVs, linkage analysis, and SNVs. With these data, a stepwise multiphasic approach to identify disease-causing variations in the family was conducted.

8p23.1, which contains a beta-defensin cluster, was detected as a region with high copy number (II-9) and low copy number (II-3 and II-7) (Figure 1-5). The defensin cluster, containing both alpha- and beta-defensins, was previously studied as a dynamic genomic region with varying copy numbers ranging from one to twelve [19]. The parents had normal copy numbers, which was in contrast to the low copy numbers seen in two children and high copy number observed in one child. A total of four haplotypes of 8p23.1 may have been inherited in this family, and each parent may have had both under- and over-amplified alleles of 8p23.1. The overall copy number of a parent can appear to be normal due to compensation of copy number from over- to under-amplified alleles [20]. In the family in this study, GSTM1 and UGT2B17, genes with frequently reported deletions [21], as well as BNTL3 and LILRB2, exhibited CNVs. I used Fisher's exact test on the affected and unaffected family members after validating this method for 8p23.1 and GSTM1 groups to determine the amplification or deletion of multiple exons

that matched the co-segregation pattern of the disease. Multiple statistically significant peaks at 8p23.1 and GSTM1 were identified, and were identical to plots from the first approach. However, there was only one statistically significant peak identified by testing the two groups that segregated with the disease, and this peak did not correlate with disease status. Thus, while WES may provide a method to identify CNV regions with highly similar sequences, determining accurate copy numbers can prove difficult.

Linkage analysis was performed to narrow down the number of candidates based on WES data. Importantly, linkage analysis with a relatively small number of markers still provides useful information. Fewer markers from WES data are available and can be obtained from a SNP microarray, and the markers that are identified may not be evenly distributed. Given these limitations, it is necessary to consider the potential disadvantages of this approach. Because I analyzed only exonic SNPs (~1% of genome-wide SNPs), critical information located outside of exons may have been lost. In addition, potential genotyping errors in linkage analyses can reduce statistical power for detecting linkage peaks or result in false positive linkage peaks [22]. Even so, the results obtained from the different data sets in this study confirmed the validity of our approach. Linkage analysis requires a large number of subjects to help identify putative loci. Unless a proper number of subjects are available, an informative result is difficult to obtain.

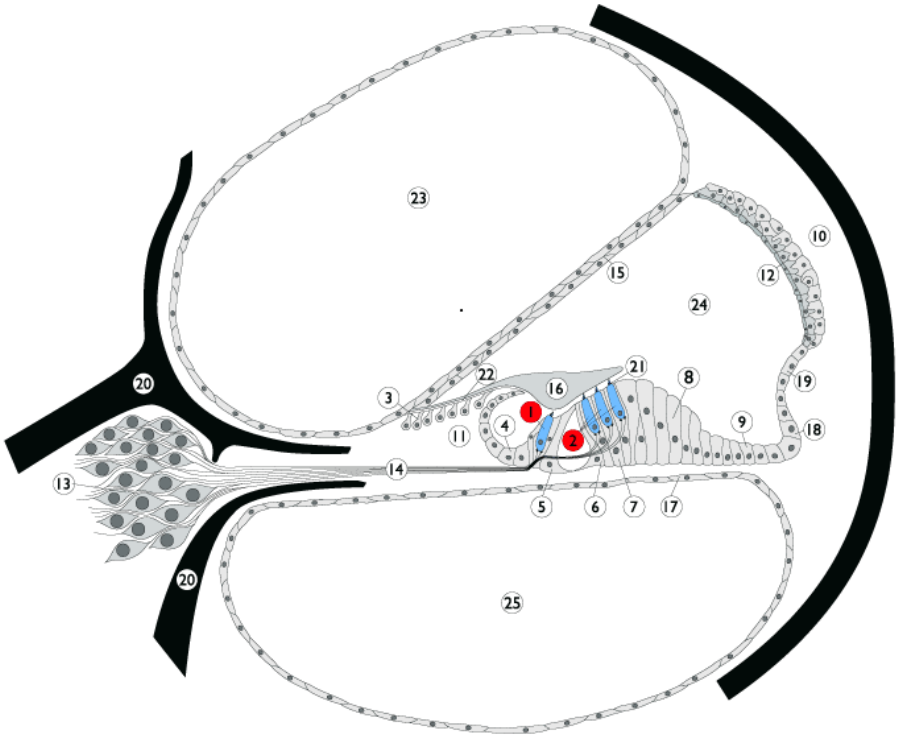
After applying linkage analysis results, the co-segregated variants were all found to be located in the loci of high LOD scores. However, linkage analysis can decrease the number of candidate variants, particularly in

instances where candidate variants are widely distributed. Additional linkage analysis of WES data demonstrated a similar performance to that of SNP microarray data and simultaneously generated results during variant calling. Considering that CNVs could be also detected using this approach, the multiphasic analysis of WES data efficiently narrowed and identified candidate variants and was advantageous compared with established methods such as initial aCGH, variant calling according to WES data alone, or linkage analysis based on SNP microarray data.

Actin is a highly conserved cytoskeletal protein that plays important roles in eukaryotic cell processes such as cell division, migration, endocytosis, and contractility. Actin isoforms are classified into two groups based on expression patterns. ACTA1, ACTA2, ACTC, and ACTG2 are “muscle” actins, predominantly expressed in striated or smooth muscle, whereas ACTB and ACTG1 are cytoplasmic “non-muscle” actins [23]. Autosomal dominant progressive sensorineural hearing loss, DFNA20/26 (MIM: 604717), is caused by a mutation in the gamma-actin gene on chromosome 17 at q25.3. Some ACTG1 mutations are associated with Baraitser-Winter syndrome, which is characterized by developmental delay, facial dysmorphologies, brain malformations, colobomas, and variable hearing loss. The constellation of these abnormalities is suggested as the most severe phenotype of ACTG1 mutations [24,25]. A genome-wide screen of DFNA20 localized candidates to 17q25.3 [26] and mapped the causative missense mutations to highly conserved actin domains of the gamma-actin gene (ACTG1) [27,28]. Tissue expression of ACTG1 in the inner ear was distributed in inner and outer hair

cells (Figure 1-13). In vivo and in vitro studies of ACTG1 indicate that it is required for reinforcement and long-term stability of actin filamentous structures of stereocilia, but not for auditory hair cell development, which is in line with the progressive nature of hearing loss related to ACTG1 mutations in humans [29,30]. Further, missense mutations in either ACTB or ACTG1 have recently been reported to cause Baraitser-Winter syndrome. Interestingly, of the 11 mutations that cause DFNA20 [27,28,31-34] and 6 mutations that cause Baraitser-Winter syndrome (see OMIM entry - *102560) that have been reported, are all missense mutations. The predicted interaction between Met305 and ATP in bovine beta-actin, a protein with a 99% identity to ACTG1, implies that the mutation of Met305 may influence ATP binding of ACTG1, which is essential for polymerization of G-actin to F-actin.

ACTG1 is predominantly expressed in intestinal epithelial and auditory hair cells [35]. Detection of exclusively missense mutations in this gene may imply that truncating mutations have more severe effects and might cause embryonic lethality. The hearing impaired subjects in this study (II-2, II-3, and II-5) did not report any gastrointestinal complaints. The subjects in this study required cochlear implants, recapitulating what has previously been reported regarding the management of patients with mutations in ACTG1 and resultant NSHL [31]. The severe phenotype and rapid progression of hearing loss to a profound level within one or two decades associated with mutations in ACTG1 necessitates an early molecular genetic diagnosis and timely auditory rehabilitation.



1. Inner hair cell 2. Outer hair cell 3. Interdental cells 4. Innter sulcus cells 5. Inner pillar cells 6. Outer pillar cells 7.Deiter's cells 8. Hensen cells 9. Claudius cells 10. Spiral ligament 11. Spiral limbus 12. Stria vascularis 13. Spiral ganglion 14. Auditory nerve 15. Reissner's membrane 16. Tectorial membrane 17. Basilar membrane 18. External sulcus cells 19. Spiral prominence 20.Bony spiral lamina 21. Reticular lamina 22. Between IDC and TM 23. Scala vestibule 24. Scala media 25. Scala tympani

Figure 1-13. **Tissue expression of ACTG1in the inner ear** (downloaded from http://hereditaryhearingloss.org/). Inner and outer hair cells have high expression of ACTG1.

Two or more platforms (aCGH, SNP array, and WES) have previously been required to generate complex genetic information such as CNVs, linkages, SNVs, and indels. Array technologies emerged several years earlier than NGS technologies and have been core tools for genetics and genomics. Based on hybridization of DNA fragments and resulted fluorescence signals, many platforms have been developed – arrayCGH, RNA expression microarray, SNPchip, etc. Since the advent of NGS technology, the two have been in competition for accuracy and efficiency. NGS is now less expensive than past and tools for NGS are steadily being developed. In the genetic research of rare diseases, arrayCGH or SNPchip is still effective for detection or narrowing the candidates. However, in general studies of Mendelian disorders, WES has primarily been utilized to obtain only SNVs and indels. In that there is more information in WES data, as many analytical methods as possible should be considered for utilization. My study agrees well with other work demonstrating that analysis of WES data also allows for CNV and linkage determination due to its quantitative traits. Given the robust nature of WES data, it is clear that the full capabilities of this relatively new technology have not yet been fully realized. Our multiphasic WES analysis proved very powerful for the interpretation and narrowing of WES results, in particular when a large amount of family data is available.

CHAPTER 2

Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss

INTRODUCTION

Targeted resequencing with target enrichment by capturing partial sequences of the genome has been developed and now enables us to endeavor genetics on the cassettes of genes that we are interested in. Custom-designed probes for capturing are used for target enrichment and large number of samples for experiments is beneficial with common interested regions. Thus, if we use this method routinely in the research or clinics, the efficiency and cost may decrease. Besides, the efficacy in detection of variants is much higher than the general WES. This is because of the limited list of target genes related to phenotype, higher production in read depth, and more availability to interpretation.

Hearing mechanism is very intricate, complex and coordinated. So, genetic hearing loss is derived from heterogeneous genes with different functional groups. >100 loci and ~50 genes have been identified as causal for hearing loss and this heterogeneity tells us again the complexity of biology and the reason of high occurrence of hearing loss. The conventional method to detect the locus and the mutation was linkage analysis and Sanger sequencing, respectively. Or if the prior studies exist on the causative gene, which describe the relation between the gene and phenotype, we can rely on the patient's phenotype to narrow the candidate genes and genotype more easily. But this case comprises only 10~20% of the entire patients and it is necessary to find some new tools for detection of variants in the remaining patients. NGS technology has been used for genetics and genomics with high performance in

many biological and clinical researches since its emergence. Due to high occurrence of monogenic mutations and heterogeneity of causative genes in hearing loss, target capturing followed by NGS technology may be the optimal choice for experiment and analysis.

Here I report the new diagnostic pipeline combining Sanger sequencing and targeted resequencing to find mutations in familial NSHL cases. Screening mutations in all exons in 80 reported deafness genes could detect candidate mutations in 13 (65%) out of 20 familial NSHL cases. Together with Sanger sequencing against four NSHL genes, the mutation detection rate was increased to 78.1% (25/32) (Figure 1-1)

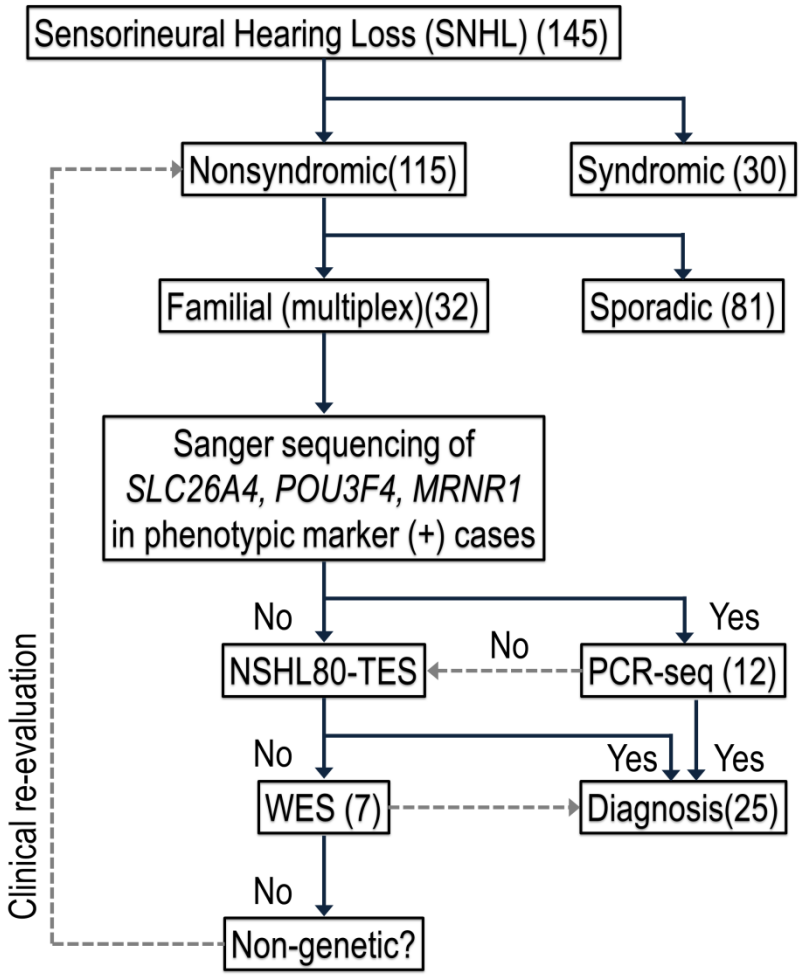


Figure 2-1. Proposed decision procedure for the genetic diagnosis of familial NSHL. 145 sensorineural hearing loss patients were recruited. Among 115 NSHL cases, 32 familial NSHL were selected because it was able to check the inheritance patterns in the family. First, 12 cases with typical clinical features by PCR Sanger sequencing were excluded. In the remaining 20 familial NSHL probands, I found candidate SNVs in 13 probands.

MATERIALS AND METHODS

Ethics statement

This study was approved by the Institutional Review Boards (IRBs) at Seoul National University Bundang Hospital (IRB-B-1007-105-402 and IRB-B-1111-139-015) and Seoul National University Hospital (IRBY-H-0905-041-281). Written informed consent from all the participants was obtained in this study. In case of children participants, the written informed consent was obtained from the parents or guardians on behalf of them.

Patient selection

Among 145 hearing impaired probands who visited our tertiary referral center and who were willing to participate the genetic test from May 2010 through April 2012, 30 probands with syndromic features were excluded. Among the remaining 115 probands, 31 families with at least two or more hearing-impaired members without any syndromic feature (multiplex families) were selected and blood samples were taken. Medical histories were collected including age at onset of hearing loss, degree and progression of hearing impairment, and other relevant clinical manifestations.

DNA preparation, Sanger sequencing and targeted resequencing

Genomic DNAs were extracted from peripheral blood as described previously [6]. Sanger sequencing was performed using specific primers for each exon as described (Table 2-1). Targeted exome sequencing was done by Otogenetics (Norcross, GA). Briefly, genomic DNA was used for NimbleGen capture methods (Roche NimbleGen Inc., Madison, WI) against 80 known deafness genes. An additional 50 bp of flanking intronic sequence were added to each exon and genomic intervals were merged using Galaxy software (http://galaxy.psu.edu). In total, I targeted 1,254 regions comprising 421,741 bp using NimbleGen methods.

Alignment, coverage calculation and variant detection

Reads were aligned to UCSC hg19 reference genome using BWA-0.6.1 with default settings [6]. To process sam/bam files and mark duplicates, Samtools and Picard were used. Local realignment around indels and base quality score recalibration were done for each samples and variants were called by unified genotyper in GATK-1.3. Perl script and Annovar were used to annotate variants and search the known SNPs and indels from dbsnp135 and 1000 genome draft. Coverages were calculated by GATK.

Table 2-1. Primer sequences used for PCR-Sanger sequencing.

Gene/variant	Exon	Forward primer	Reverse primer	Amplicon (bp)
COCH	Exon 4	GATGCCCTGAAAAAGTGTGG	TCACAGGTTTTTCCATCAAGG	289
	Exon 3	AAAACAACCTTGTGGCTTGC	CCAGATGGGTAAAGCAGGAA	408
COL11A2	Exon 30	CCCATCCTGACCCCAGTG	CACTGTTGCCCATTTCTCCT	249
EYA4	Exon 11	GCCATCAGGAGGTTTCTATTGTAT	GTAGGCATACCTCCAGGTCACTAT	259
GJB2	Exon 2	TCCTAGCTAGTGATTCCTGTGTTG	AGCCGTCGTACATGACATAGAAG	547
GJB3	Exon 2	GATGAGCAGAAGGACTTTGACTG	AACTCAATGATGAGCTTGAAGATG	293
MYO3A	Exon 7	GCAATTGAAAGCTCTTTATATGAGT	AAAAGGAAAGTCAAACAGGATCA G	199
	Exon 16	ATATCAGCACTCACAGTCGTTGTT	TAAAATTAAGGGAAAAGTGAAA	263
MYO6	Exon 8	TATTTTGTAATGTTCCGTCATGCT	TCCTGCAACCATCTAAAGTAACAA	286
OTOF	Exon 24	TCAGTCCCTCCCATGCAG	TGTGCTGACCCCAAGACC	244
	Exon 8	GGTGATCACACCTGTCCCTTA	CAACTCCCAGCCTCCAGTC	327
OTOR	Exon 2	GGGAATTATCAGTCACTCTGATTTT	TGAAGGTAGGTAATCAAGGAAAGG	314
STRC	Exon 20	CTCCAGTCTCAGGGGAAGTG	TCTGAAGTTCTCGAAGGTCCA	208
WFS1	Exon 8	ATCGACTTCTTCGCCTTCTTC	CTTGAATTGGCCCTACCTGAAG	435
MTRNR1	mitochondria	TCTACCCCAGAAAACTACGATAGC	TGTTAAGCTACACTCTGGTTCGTC	248

Table 2-2. List of 80 genes related to NSHL for targeted resequencing.

Symbol	Annotation	HGNC	OMIM
ACTB	actin, beta	132	[607371] DYSTONIA, JUVENILE-ONSET
ACTG1	actin, gamma 1	144	[604717] DEAFNESS, AUTOSOMAL DOMINANT 20; DFNA 20
ATP6V1B1	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B1	853	[267300] RENAL TUBULAR ACIDOSIS, DISTAL, WITH PR OGRESSIVE NERVE DEAFNESS
BCS1L	BCS1-like (S. cerevisiae)	1020	[603358] GRACILE SYNDROME [262000] BJORNSTAD SYNDROME; BJS [256000] LEIGH SYNDROME; LS [124000] MITOCHONDRIAL COMPLEX III DEFICIENCY
BSND	Bartter syndrome, infantile, with sensorineural deafness (Barttin)	16512	[602522] BARTTER SYNDROME, TYPE 4A
CATSPER 2	cation channel, sperm associated 2	18810	[611102] DEAFNESS, SENSORINEURAL, AND MALE INFE RTILITY
CCDC50	coiled-coil domain containing 50	18111	[607453] DEAFNESS, AUTOSOMAL DOMINANT 44; DFNA 44
CDH23	cadherin-related 23	13733	[601386] DEAFNESS, AUTOSOMAL RECESSIVE 12; DFNB 12 [601067] USHER SYNDROME, TYPE ID; USH1D [276900] USHER SYNDROME, TYPE I; USH1
CLDN14	claudin 14	2035	[614035] DEAFNESS, AUTOSOMAL RECESSIVE 29; DFNB 29
COCH	coagulation factor C homolog, cochlin (Limulu s polyphemus)	2180	[601369] DEAFNESS, AUTOSOMAL DOMINANT 9; DFNA9

COL11A2	collagen, type XI, alpha 2	2187	[614524] FIBROCHONDROGENESIS 2; FBCG2 [609706] DEAFNESS, AUTOSOMAL RECESSIVE 53; DFNB 53 [601868] DEAFNESS, AUTOSOMAL DOMINANT 13; DFNA 13 [277610] WEISSENBACHER-ZWEYMULLER SYNDROME; WZS [215150] OTOSPONDYLOMEGAEPIPHYSEAL DYSPLASIA; OSMED [184840] STICKLER SYNDROME, TYPE III; STL3
COL9A3	collagen, type IX, alpha 3	2219	[603932] INTERVERTEBRAL DISC DISEASE; IDD [600969] EPIPHYSEAL DYSPLASIA, MULTIPLE, 3; EDM3
CRYM	crystallin, mu	2418	#N/A
DFNA5	deafness, autosomal dominant 5	2810	[600994] DEAFNESS, AUTOSOMAL DOMINANT 5; DFNA5
DFNB31	deafness, autosomal recessive 31	16361	[611383] USHER SYNDROME, TYPE IID; USH2D [607084] DEAFNESS, AUTOSOMAL RECESSIVE 31; DFNB 31
DFNB59	deafness, autosomal recessive 59	29502	[610220] DEAFNESS, AUTOSOMAL RECESSIVE 59; DFNB 59
DIAPH1	diaphanous homolog 1 (Drosophila)	2876	[124900] DEAFNESS, AUTOSOMAL DOMINANT 1; DFNA1
DSPP	dentin sialophosphoprotein	3054	[605594] DEAFNESS, AUTOSOMAL DOMINANT 39, WITH DENTINOGENESIS IMPERFECTA 1 [125500] DENTINOGENESIS IMPERFECTA, SHIELDS TYPE III [125490] DENTINOGENESIS IMPERFECTA 1; DGI1 [125420] DENTIN DYSPLASIA, TYPE II
ERCC2	excision repair cross-complementing rodent repair deficiency, complementation group 2	3434	[610756] CEREBROOCULOFACIOSKELETAL SYNDROME 2; COFS2 [601675] TRICHOTHIODYSTROPHY, PHOTOSENSITIVE;

			TTDP [278730] XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP D; XPD
ERCC3	excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	3435	[610651] XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP B; XPB [601675] TRICHOTHIODYSTROPHY, PHOTOSENSITIVE; TTDP
ESPN	espin	13281	[609006] DEAFNESS, AUTOSOMAL RECESSIVE 36, WITH OR WITHOUT VESTIBULAR INVOLVEMENT
ESRRB	estrogen-related receptor beta	3473	[608565] DEAFNESS, AUTOSOMAL RECESSIVE 35; DFNB 35
EYA4	eyes absent homolog 4 (Drosophila)	3522	[605362] CARDIOMYOPATHY, DILATED, 1J; CMD1J [601316] DEAFNESS, AUTOSOMAL DOMINANT 10; DFNA 10
FGF3	fibroblast growth factor 3	3681	[610706] DEAFNESS, CONGENITAL, WITH INNER EAR AGENESIS, MICROTIA, AND MICRODONTIA
GATA3	GATA binding protein 3	4172	[146255] HYPOPARATHYROIDISM, SENSORINEURAL DEAFNESS, AND RENAL DISEASE; HDR
GJA1	gap junction protein, alpha 1, 43kDa	4274	[600309] ATRIOVENTRICULAR SEPTAL DEFECT 3; AVSD 3 [257850] OCULODENTODIGITAL DYSPLASIA, AUTOSOMAL RECESSIVE [241550] HYPOPLASTIC LEFT HEART SYNDROME 1; HLHS1 [234100] HALLERMANN-STREIFF SYNDROME; HSS [186100] SYNDACTYLY, TYPE III [164200] OCULODENTODIGITAL DYSPLASIA; ODDD
GJB1	gap junction protein, beta 1, 32kDa	4283	[302800] CHARCOT-MARIE-TOOTH DISEASE, X-LINKED DOMINANT, 1; CMTX1

			[145900] HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS
GJB2	gap junction protein, beta 2, 26kDa	4284	[602540] ICHTHYOSIS, HYSTRIX-LIKE, WITH DEAFNESS [601544] DEAFNESS, AUTOSOMAL DOMINANT 3A; DFNA3A [220290] DEAFNESS, AUTOSOMAL RECESSIVE 1A; DFNB1A [149200] KNUCKLE PADS, LEUKONYCHIA, AND SENSORINEURAL DEAFNESS [148350] KERATODERMA, PALMOPLANTAR, WITH DEAFNESS [148210] KERATITIS-ICHTHYOSIS-DEAFNESS SYNDROME, AUTOSOMAL DOMINANT [124500] DEAFNESS, CONGENITAL, WITH KERATOPACHYDERMIA AND CONSTRICTIONS OF
GJB3	gap junction protein, beta 3, 31kDa	4285	[612644] DEAFNESS, AUTOSOMAL DOMINANT 2B; DFNA2B [603324] GAP JUNCTION PROTEIN, BETA-3; GJB3 [600101] DEAFNESS, AUTOSOMAL DOMINANT 2A; DFNA2A [220290] DEAFNESS, AUTOSOMAL RECESSIVE 1A; DFNB1A [133200] ERYTHROKERATODERMIA VARIABILIS ET PROGRESSIVA; EKVP
GJB4	gap junction protein, beta 4, 30.3kDa	4286	[133200] ERYTHROKERATODERMIA VARIABILIS ET PROGRESSIVA; EKVP
GJB6	gap junction protein, beta 6, 30kDa	4288	[612645] DEAFNESS, AUTOSOMAL RECESSIVE 1B; DFNB1B [612643] DEAFNESS, AUTOSOMAL DOMINANT 3B; DFN

			A3B [601544] DEAFNESS, AUTOSOMAL DOMINANT 3A; DFNA3A [220290] DEAFNESS, AUTOSOMAL RECESSIVE 1A; DFNB1A [129500] ECTODERMAL DYSPLASIA, HIDROTIC, AUTOSOMAL DOMINANT
GRHL2	grainyhead-like 2 (Drosophila)	2799	[608641] DEAFNESS, AUTOSOMAL DOMINANT 28; DFNA28
GSTP1	glutathione S-transferase pi 1	4638	#N/A
JAG1	jagged 1	6188	[601920] JAGGED 1; JAG1 [187500] TETRALOGY OF FALLOT; TOF [118450] ALAGILLE SYNDROME 1; ALGS1
KCNE1	potassium voltage-gated channel, Isk-related family, member 1	6240	[613695] LONG QT SYNDROME 5; LQT5 [612347] JERVELL AND LANGE-NIELSEN SYNDROME 2; JLNS2 [220400] JERVELL AND LANGE-NIELSEN SYNDROME 1; JLNS1
KCNJ10	potassium inwardly-rectifying channel, subfamily J, member 10	6256	[612780] SEIZURES, SENSORINEURAL DEAFNESS, ATAXIA, MENTAL RETARDATION, AND [600791] ENLARGED VESTIBULAR AQUEDUCT; EVA
KCNQ4	potassium voltage-gated channel, KQT-like subfamily, member 4	6298	[600101] DEAFNESS, AUTOSOMAL DOMINANT 2A; DFNA2A
LHFPL5	lipoma HMGIC fusion partner-like 5	21253	[610265] DEAFNESS, AUTOSOMAL RECESSIVE 67; DFNB67
LHX3	LIM homeobox 3	6595	[262600] PITUITARY HORMONE DEFICIENCY, COMBINED, 2; CPHD2 [221750] PITUITARY HORMONE DEFICIENCY, COMBINED, 3; CPHD3

LRTOMT	leucine rich transmembrane and 0-methyltransferase domain containing	25033	[611451] DEAFNESS, AUTOSOMAL RECESSIVE 63; DFNB 63
MARVEL D2	MARVEL domain containing 2	26401	[610153] DEAFNESS, AUTOSOMAL RECESSIVE 49; DFNB 49
MTAP	methylthioadenosine phosphorylase	7413	#N/A
MYH14	myosin, heavy chain 14, non-muscle	23212	[614369] PERIPHERAL NEUROPATHY, MYOPATHY, HOARSENESS, AND HEARING LOSS; PNMHH [600652] DEAFNESS, AUTOSOMAL DOMINANT 4; DFNA4
MYH9	myosin, heavy chain 9, non-muscle	7579	[605249] SEBASTIAN SYNDROME; SBS [603622] DEAFNESS, AUTOSOMAL DOMINANT 17; DFNA 17 [600208] MACROTHROMBOCYTOPENIA AND PROGRESSIVE SENSORINEURAL DEAFNESS [155100] MAY-HEGGLIN ANOMALY; MHA [153650] EPSTEIN SYNDROME, [153640] FECHTNER SYNDROME; FTNS
MYO15A	myosin XVA	7594	[600316] DEAFNESS, AUTOSOMAL RECESSIVE 3; DFNB3
MYO1A	myosin IA	7595	[607841] DEAFNESS, AUTOSOMAL DOMINANT 48; DFNA 48
MYO1C	myosin IC	7597	#N/A
MYO1F	myosin IF	7600	#N/A
MYO3A	myosin IIIA	7601	[607101] DEAFNESS, AUTOSOMAL RECESSIVE 30; DFNB 30
MYO6	myosin VI	7605	[607821] DEAFNESS, AUTOSOMAL RECESSIVE 37; DFNB 37 [606346] DEAFNESS, AUTOSOMAL DOMINANT 22; DFNA 22
MYO7A	myosin VIIA	7606	[601317] DEAFNESS, AUTOSOMAL DOMINANT 11; DFNA 11

			[600060] DEAFNESS, AUTOSOMAL RECESSIVE 2; DFNB2 [276900] USHER SYNDROME, TYPE I; USH1
NR2F1	nuclear receptor subfamily 2, group F, member 1	7975	#N/A
OTOA	otoancorin	16378	[607039] DEAFNESS, AUTOSOMAL RECESSIVE 22; DFNB 22
OTOF	otoferlin	8515	[601071] DEAFNESS, AUTOSOMAL RECESSIVE 9; DFNB9
OTOR	otoraplin	8517	#N/A
PAX3	paired box 3	8617	[268220] RHABDOMYOSARCOMA 2; RMS2 [193500] WAARDENBURG SYNDROME, TYPE 1; WS1 [148820] WAARDENBURG SYNDROME, TYPE 3; WS3 [122880] CRANIOFACIAL-DEAFNESS-HAND SYNDROME; CDHS
PCDH15	protocadherin-related 15	14674	[609533] DEAFNESS, AUTOSOMAL RECESSIVE 23; DFNB 23 [602083] USHER SYNDROME, TYPE IF; USH1F [601067] USHER SYNDROME, TYPE ID; USH1D [276900] USHER SYNDROME, TYPE I; USH1
PDZD7	PDZ domain containing 7	26257	[605472] USHER SYNDROME, TYPE IIC; USH2C [276901] USHER SYNDROME, TYPE IIA; USH2A
PMP22	peripheral myelin protein 22	9118	[180800] ROUSSY-LEVY HEREDITARY AREFLEXIC DYSTASIA [162500] NEUROPATHY, HEREDITARY, WITH LIABILITY TO PRESSURE PALSIES; HNPP [145900] HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS [139393] GUILLAIN-BARRE SYNDROME, FAMILIAL; GBS [118300] CHARCOT-MARIE-

			TOOTH DISEASE AND DEAFNESS [118220] CHARCOT-MARIE- TOOTH DISEASE, DEMYELINATING, TYPE 1A; CMT1A
POU4F3	POU class 4 homeobox 3	9220	[602459] DEAFNESS, AUTOSOMAL DOMINANT 15; DFNA 15
RDX	radixin	9944	[611022] DEAFNESS, AUTOSOMAL RECESSIVE 24; DFNB 24
SLC17A8	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), me mber 8	20151	[605583] DEAFNESS, AUTOSOMAL DOMINANT 25; DFNA 25
SLC26A4	solute carrier family 26, member 4	8818	[600791] ENLARGED VESTIBULAR AQUEDUCT; EVA [274600] PENDRED SYNDROME; PDS
SLC26A5	solute carrier family 26, member 5 (prestin)	9359	[613865] DEAFNESS, AUTOSOMAL RECESSIVE 61; DFNB 61
SLC4A11	solute carrier family 4, sodium borate transport er, member 11	16438	[613268] CORNEAL DYSTROPHY, FUCHS ENDOTHELIAL , 4; FECD4 [217700] CORNEAL ENDOTHELIAL DYSTROPHY 2, AUT OSOMAL RECESSIVE; CHED2 [217400] CORNEAL DYSTROPHY AND PERCEPTIVE DEA FNESS
SOX2	SRY (sex determining region Y)-box 2	11195	[206900] MICROPHTHALMIA, SYNDROMIC 3; MCOPS3
SPINK5	serine peptidase inhibitor, Kazal type 5	15464	[256500] NETHERTON SYNDROME; NETH [147050] IgE RESPONSIVENESS, ATOPIC; IGER
STRC	stereocilin	16035	[611102] DEAFNESS, SENSORINEURAL, AND MALE INFE RTILITY [603720] DEAFNESS, AUTOSOMAL RECESSIVE 16; DFNB 16
TBL1X	transducin (beta)-like 1X-linked	11585	#N/A
TCF21	transcription factor 21	11632	#N/A

TECTA	tectorin alpha	11720	[603629] DEAFNESS, AUTOSOMAL RECESSIVE 21; DFNB 21 [601543] DEAFNESS, AUTOSOMAL DOMINANT 12; DFNA 12
TIMM8A	translocase of inner mitochondrial membrane 8 homolog A (yeast)	11817	[311150] OPTICOACOUSTIC NERVE ATROPHY WITH DE MENTIA [304700] MOHR-TRANEBJAERG SYNDROME; MTS
TMC1	transmembrane channel-like 1	16513	[606705] DEAFNESS, AUTOSOMAL DOMINANT 36; DFNA 36 [600974] DEAFNESS, AUTOSOMAL RECESSIVE 7; DFNB7
TMIE	transmembrane inner ear	30800	[600971] DEAFNESS, AUTOSOMAL RECESSIVE 6; DFNB6
TMPRSS3	transmembrane protease, serine 3	11877	[605316] DEAFNESS, AUTOSOMAL RECESSIVE 10; DFNB 10 [601072] DEAFNESS, CHILDHOOD-ONSET NEUROSENSORY, AUTOSOMAL RECESSIVE 8; D FNB8
TMPRSS5	transmembrane protease, serine 5	14908	#N/A
TRIOBP	TRIO and F-actin binding protein	17009	[609823] DEAFNESS, AUTOSOMAL RECESSIVE 28; DFNB 28
USH1C	Usher syndrome 1C (autosomal recessive, sever e)	12597	[602092] DEAFNESS, AUTOSOMAL RECESSIVE 18; DFNB 18 [276904] USHER SYNDROME, TYPE IC; USH1C [276900] USHER SYNDROME, TYPE I; USH1
WFS1	Wolfram syndrome 1 (wolframin)	12762	[614296] WOLFRAM-LIKE SYNDROME, AUTOSOMAL DOMINANT; WFSL [600965] DEAFNESS, AUTOSOMAL DOMINANT 6; DFNA6 [222300] WOLFRAM SYNDROME 1; WFS1 [125853] DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM

Model of independent uncaptured exons

To evaluate the correlation of capture performance between each samples, I compared experimental and expected distribution in number of exons that were commonly uncaptured within from 0 to 20 samples. Expectation number was calculated by the model assuming that uncapturing of exons occur independently between samples. Binomial model was not used because the difference of the numbers of uncaptured exons was not ignorable. I defined that uncaptured exon is an exon within which % of bases above 10, 50 or 100 of read depth is less than 1%.

p_k: ratio of uncaptured exons in kth sample

P(#=n): probability that number of samples having uncaptured exons in common is n

Here, the number of samples is 20 and the number of exons is 1254.

$$\mathrm{P}(\#=\mathrm{n}) = \sum p_1^{i_1} \cdots p_{20}^{i_{20}} (1-p_1)^{j_1} \cdots (1-p_{20})^{j_{20}}$$

(sum for all combinations where, $\sum_{k=1}^{20} i_k = n, \sum_{k=1}^{20} j_k = 20 - n$ ($0 \leq n \leq 20$) for i_k and j_k= 0 or 1)

Due to too large number of combinations, each probability(P) was calculated with permutation 1,000 times with Python using a module ‘‘decimal’’ for precision, instead of summing all the combinations (but P(#= 0) was calculated directly.). Then, expected counts were obtained.

$$\mathrm{E}(\# = \mathrm{n}) = \mathrm{P}(\# = \mathrm{n}) \times 1254$$

RESULTS

145 sensorineural hearing loss cases have been collected for a molecular genetic diagnosis in SNUH and SNUBH. Among them, 32 multiplex familial cases were selected to find genetic variations for diagnosis and genetic counseling because they were available to utilize co-segregation information in the family to validate the causative mutation. The new diagnostic pipeline combining PCR and targeted resequencing was established. Eleven cases showed clearly defined phenotype related to the mutations in SLC26A4, POU3F4 or mitochondrial DNA genes (Table 2-3 and Figure 2-2). Temporal bone CT was taken to rule out any abnormality of the inner ear. Cases with characteristic radiologic markers such as bilateral enlarged vestibular aqueduct (5 probands) or incomplete partition type III (5 probands) were directly subject to Sanger sequencing of the corresponding candidate genes, SLC26A4 and POU3F4, respectively. Mitochondrial DNA was sequenced for one family that showed characteristic maternal inheritance of hearing loss. In these eleven families, mutations could be found successfully by PCR sequencing, which were mostly located in the reported sites (Table 2-3). GJB2 sequencing was performed for the remaining 21 hearing impaired probands because the mutation in GJB2 was most frequent among familial NSHL cases. I found two cases (SJ19-19 and SH35-75) with known pathogenic mutations in GJB2 gene.

Table 2-3. Mutations of SLC26A4, POU3F4, GJB2 and MTRNR1 in 12 familial NSHL found by PCR-Sanger sequencing.

Patient	Characteristic phenotype	Gene	Mutat-ion type[#]	RefSeq ID	Chr	Exon	Nucleotide	Protein	MAF[*]	dbSNP135
SB02-6	Incomplete partition type III	POU3F4	mis.	NM_000307	X	exon 1	c.686A>G	p.Gln229Arg	-	-
SB07-18	Incomplete partition type III	POU3F4	Fs. del.	NM_000307	X	exon 1	c.1060delA	p.Thr354GlnfsX115	-	-
SB08-19	Incomplete partition type III	POU3F4	Fs. ins.	NM_000307	X	exon 1	c.950dupT	p.Leu317PhefsX12	-	-
SB09-21	Incomplete partition type III	POU3F4	mis.	NM_000307	X	exon 1	c.632C>T	p.Thr211Met	-	-
SB13-29	Incomplete partition type III	POU3F4	ns.	NM_000307	X	exon 1	c.623T>A	p.Leu208X	-	-
SB16-34	Nonsyndromic EVA	SLC26A4	mis.	NM_000441	7	exon 19	c.A2168G	p.H723R	0.001	rs121908362
SB23-54	Nonsyndromic EVA	SLC26A4	mis.	NM_000441	7	exon 19	c.A2168G	p.H723R	0.001	rs121908362
SB28-61	Nonsyndromic EVA	SLC26A4	mis.	NM_000441	7	exon 19	c.A2168G	p.H723R	0.001	rs121908362
SJ07-7	Nonsyndromic EVA	SLC26A4	mis.	NM_000441	7	exon 19	c.A2168G	p.H723R	0.001	rs121908362
SJ20-20	Nonsyndromic EVA	SLC26A4	mis.	NM_000441	7	exon 19	c.A2168G	p.H723R	0.001	rs121908362
SH07-19	Maternal transmission	MTRNR1	mis.		Mt		1,555A>G		-	-
SJ19-19	no specific phenotype	GJB2	Fs. del.	NM_004004	13	exon 2	c.299_300del	p.H100RfsX14	-	-
			Fs. del.	NM_004004	13	exon 2	c.235delC	p.L79CfsX3	-	-

#mis.:missense, ns.:nonsense, Fs.del.:frameshift deletion, Fs.ins.:frameshift insertion. *MAF:minor allele frequency from 1,000 Genome.

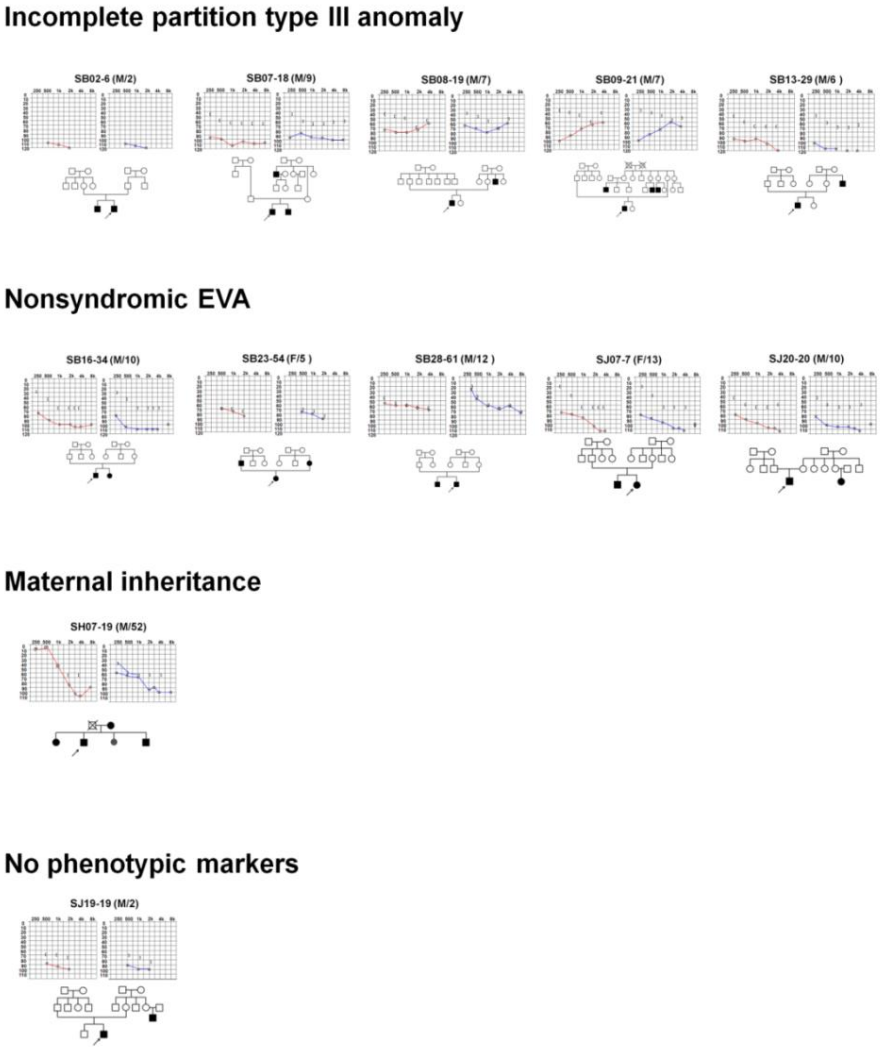


Figure 2-2. Audiogram and pedigree for 12 familial NSHL patients used in PCR-Sanger sequencing.

Next, I applied targeted resequencing with NGS technology to 20 probands of the remaining familial NSHL cases including one GJB2 positive multiplex family (SH35) to screen all 80 reported NSHL-related genes. I have captured 1254 exons of 80 genes (Table 2-2) spanning ~480 kb in 20 probands from multiplex families for targeted exome sequencing. Mean read depth in 20 cases was 218.2±56.1 and 88.9±3.7% of bases was read in more than x10 coverage (Table 2-4). About 90% exons in all patients were captured with ≥99% of bases at ≥10 of read depth. Missed or low-coverage exons were shared between samples, though different experimental procedures shared different uncaptured exons (Figure 2-3). This ensures that most of captured exons were shared through samples, which does not disturb the following analysis of variant detection. The fraction of well-captured exons was much more than expectation by the model of independent uncaptured exons (Figure 2-4).

Table 2-4. Qualities of targeted resequencing in 20 familial NSHL.

Sample	Depth							SNP				Indel
	Mean	Q3	Median	Q1	%≥10	%≥50	%≥100	Total	synon	nonsyn	stopgain	Total
SB14-30	182.0	253	154	62	88.2	77.4	65.4	176	100	75	1	9
SB38-75	309.6	478	228	60	86.3	77.0	67.9	204	118	85	1	14
SB40-77	348.8	500	266	65	87.3	77.8	70.0	181	94	87	0	17
SB4-11	283.6	434	204	52	86.1	75.5	66.5	190	101	87	2	16
SB41-78	331.2	500	247	63	86.6	77.2	69.0	189	108	79	2	12
SB47-91	232.2	342	163	43	84.7	72.7	62.4	203	116	86	1	18
SB50-94	212.6	313	147	34	84.5	71.0	59.6	186	99	87	0	19
SB54-101	200.7	258	171	102	93.4	88.2	75.5	214	136	77	1	7
SB55-102	194.1	250	163	97	93.3	88.0	74.0	156	87	69	0	10
SB60-107	211.6	272	177	105	93.5	88.8	76.5	213	123	89	1	12
SB61-109	189.2	242	161	96	93.3	87.7	73.3	232	140	92	0	15
SH10-28	211.1	319	148	37	85.0	71.2	60.2	198	107	91	0	12
SH14-37	192.7	268	164	67	88.0	78.4	67.1	212	128	84	0	12
SH20-47	161.6	223	136	55	88.2	76.1	62.1	224	133	91	0	11
SH21-50	216.1	303	186	75	89.0	79.5	70.5	195	105	88	2	11
SH23-52	203.3	285	173	69	88.2	78.7	68.4	227	130	97	0	11
SH27-61	172.4	222	144	85	93.2	86.9	68.7	222	126	96	0	18
SH35-75	144.1	209	95	24	82.2	64.7	48.6	221	129	92	0	18
SH40-89	181.9	235	150	90	93.2	87.4	70.7	235	143	92	0	14
SH41-90	185.9	239	156	92	93.1	87.3	71.9	237	141	96	0	13
Average	218.2	307.3	171.7	68.7	88.9	79.6	67.4	205.8	118.2	87.0	0.6	13.5
SD	56.1	94.9	39.4	23.8	3.7	7	6.5	21.8	17.1	7.3	0.8	3.4

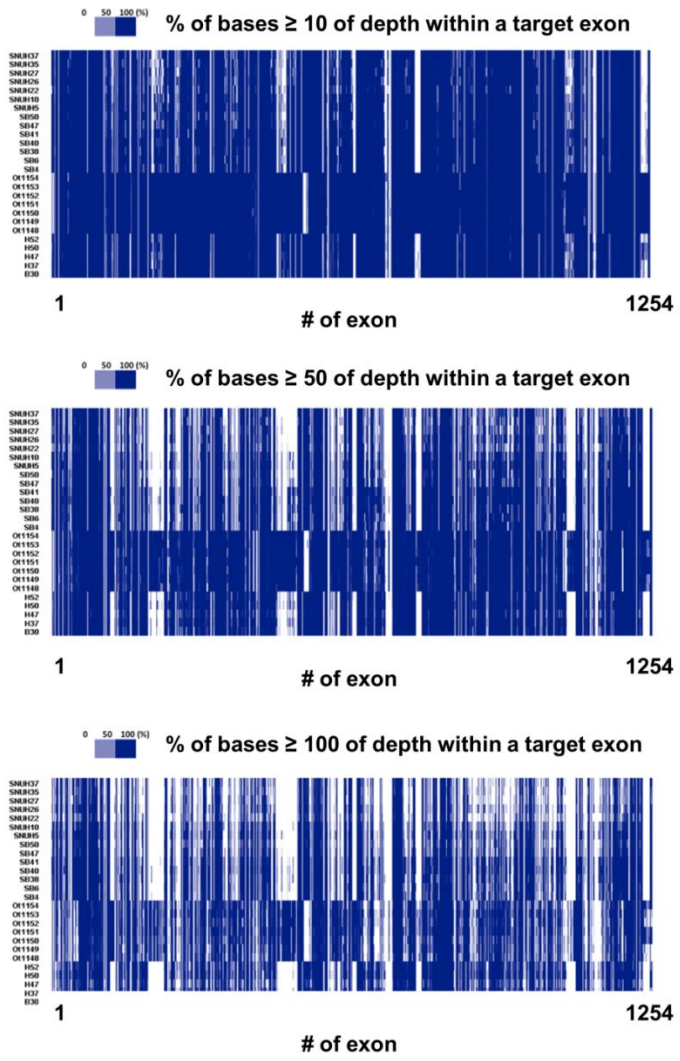


Figure 2-3. Heatmap for the performance of target capture. Heatmap was drawn for percentage of bases ≥ depth 10, 50 or 100 within all target exons and samples. Most exons were uncaptured in common samples, and samples were grouped by the common uncaptured exons.

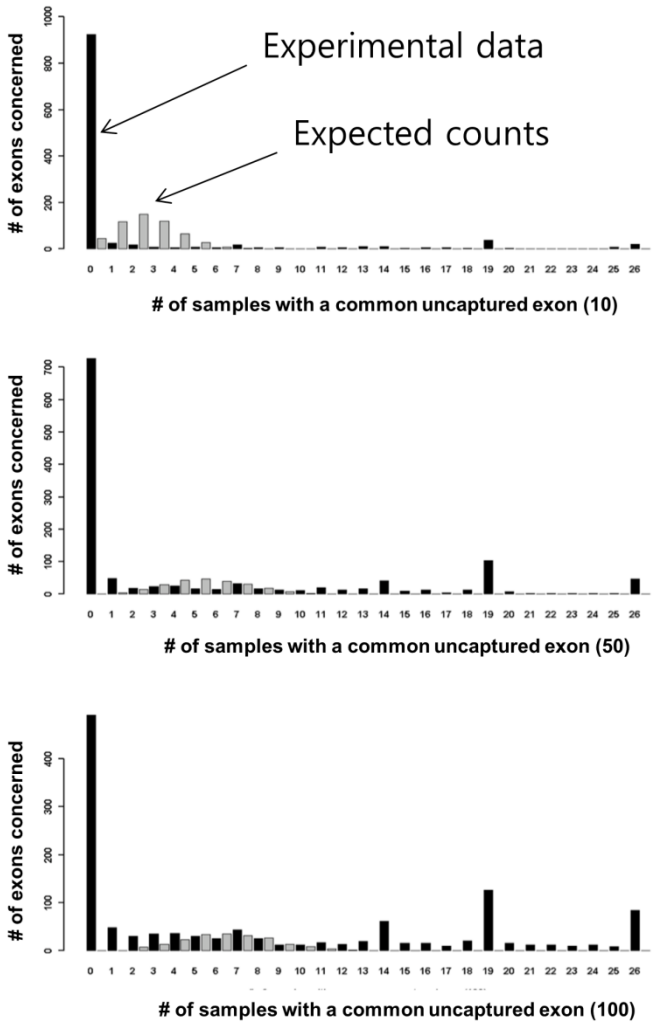


Figure 2-4. Barplot for performance of target capture. Barplot was drawn for comparison between experimental and expected distribution of number of exons that were commonly uncaptured from 0 to 20 samples. Note that the exons in n=0 were captured in all 20 samples, and at n=0, the experimental counts are greater than the expected ones due to common uncaptured exons.

Rare single nucleotide variations (SNV) or indels of which MAF (minor allele frequency) are <0.01 in 1,000 Genome Project were selected, following five steps of filtering to find candidate mutations related to hearing loss in each patient (Table 2-5, Figure 2-5 and 2-6). In the first basic filtering step, variations with a quality score of less than 20 were discarded, and for heterozygous alleles, only the alleles with a ratio (coverage of variant over the total coverage) of 20% or more were included. The average number of variants was 4.8±0.42 per patient after basic filtering. As the second step, I checked inheritance pattern of multiplex family of each proband, and excluded the variants which were not matched with the patient's inheritance pattern. Autosomal dominant pattern was matched to heterozygous variants and autosomal recessive pattern was matched to homozygous or compound heterozygous variants, respectively. According to the information on the inheritance, I could significantly reduce the average numbers of candidate mutations to 1.95±0.29 per patient (t-test p= 2.861026). In three families, all the mutations were not matched with the inheritance pattern. In the third step, 39 variants from the 18 probands by Sanger sequencing were validated and 36 variants (92.3%) were confirmed (Figure 2-7 and 2-8). When 160 normal hearing control subjects were checked for the variants by Sanger sequencing, seven variants were also found in Korean population as the fourth step. As the final step for the filtering, segregation and/or phenotype were investigated whether they matched to confirm the causality of the variant for deafness. Nineteen variants were examined in nine families by Sanger sequencing in all the family members to exclude 8 variants. The patient's audiogram was also

examined to match the candidate genes with the patient's phenotype and eleven variants were ruled out, too. Especially, in cases where the segregation study could not be performed, the audiogram configuration was relied upon. Molecular genetic diagnosis was made in four subjects (SB61-109, SB55-102, SB50-94 and SB47-91) despite the lack of segregation study results, since their audiograms were well matched the previously reported characteristic audiogram configuration. Finally, I was able to find most likely causative mutations in 13 out of 20 multiplex hearing loss families (Table 2-6).

Table 2-5. Number of candidate SNVs in 20 familial NSHL through five filtering steps.

Patient	1) basic filtering	2)inheritance pattern	3) Sanger sequencing	4) Control	5) Clinical feature		Final
					Segregation	Audiogram profile matching	
ADNSHL							
SB14-30	8	1	1	1	-	1	1
SB40-77	6	3	3	2	-	0	0
SB41-78	3	1	1	0	-	-	0
SB50-94	3	2	1	1	-	1	1
SB54-101	5	4	4	3	1	1	1
SB55-102	1	0	-	-	-	0	0
SB60-107	6	3	2	2	1	1	1
SB61-109	4	2	2	1	-	1	1
SH14-37	4	2	2	1	1	1	1
SH20-47	5	1	1	1	1	1	1
SH21-50	4	3	3	2	1	1	1
SH40-89	6	5	5	5	2	1	1
SH41-90	2	1	1	1	0	0	0
ARNSHL							
SB04-11	8	2	2	2	2	2	2
SB38-75	4	2	2	2	2	2	2
SB47-91	5	3	3	2	-	2	2
SH10-28	5	2	1	1	-	-	0
SH23-52	3	0	0	-	-	-	0
SH27-61	7	0	0	0	-	-	0
SH35-75	7	2	2	2	-	?	2

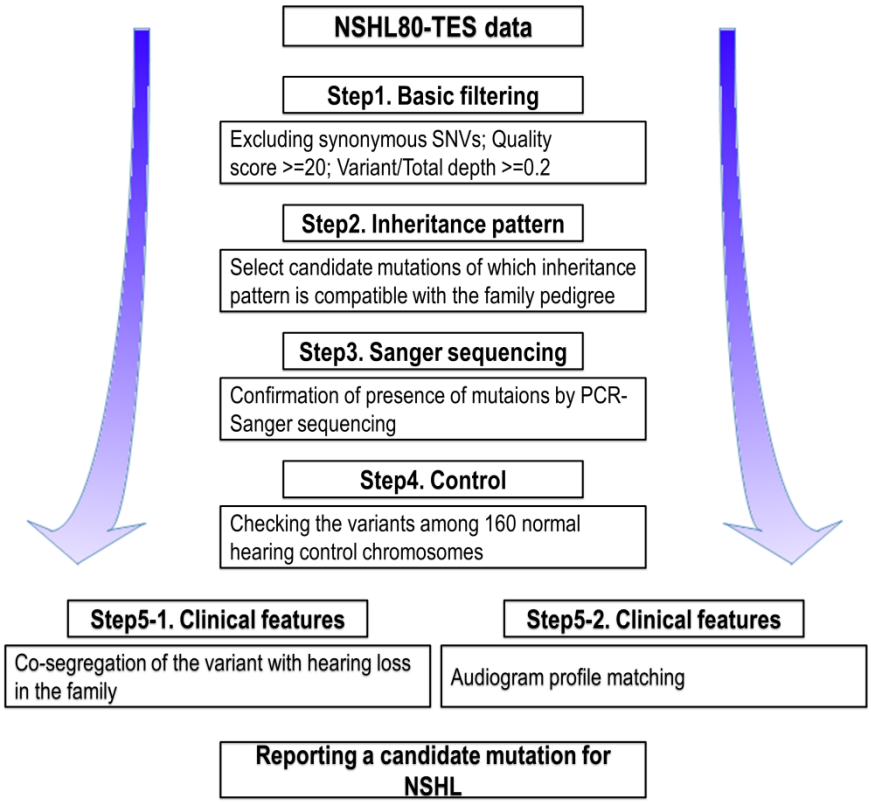


Figure 2-5. Analysis flow of NSHL-80 targeted resequencing on familial NSHL. Targeted resequencing data from 20 familial NSHL cases were filtered through five steps to select candidate SNVs in NSHL genes.

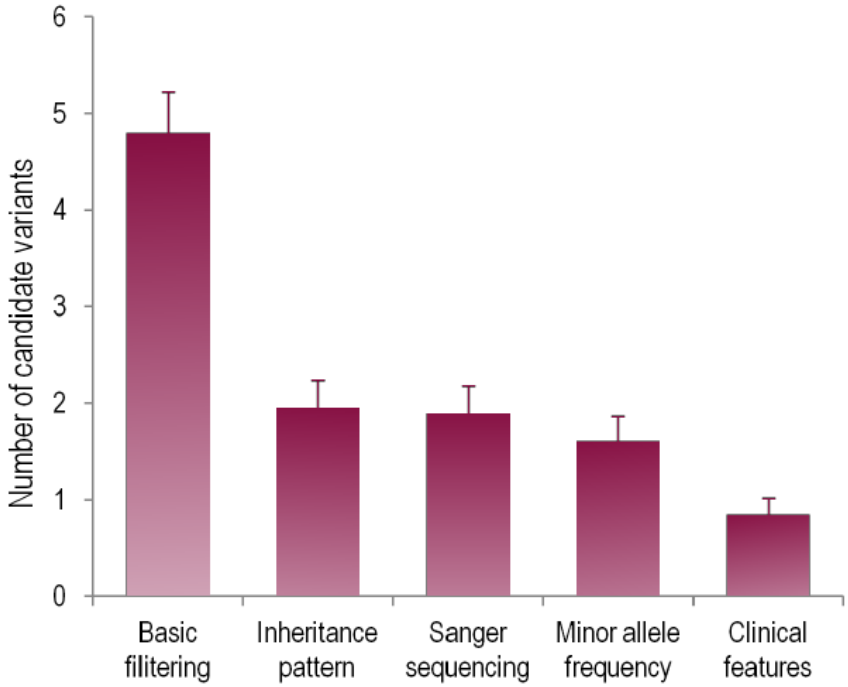


Figure 2-6. **Average number of candidate SNVs with five filtering steps**. An average number of candidate SNVs with standard errors were shown at five filtering steps.

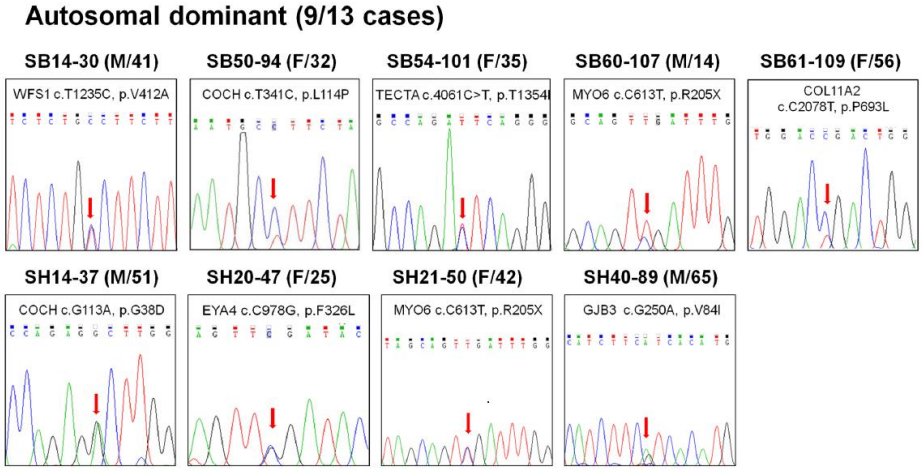


Figure 2-7. **Validation of candidate mutations in AD NSHL by Sanger sequencing**. Candidate mutations in 9 autosomal dominant NSHL families were shown in chromatogram of Sanger sequencing.

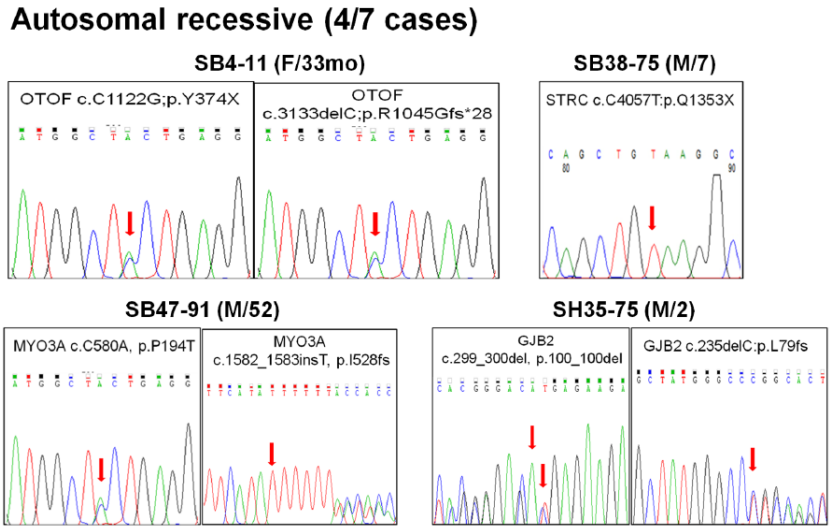


Figure 2-8. **Validation of candidate mutations in AR NSHL by Sanger sequencing**. Candidate mutations in 4 autosomal recessive NSHL families were shown in chromatogram of Sanger sequencing.

Table 2-6. List of final candidate SNVs in 13 familial NSHL.

Patient	Gene	Type	RefSeq ID	Chr	Exon	Nucleotide	Protein	Cov. of Ref	Cov. of Var	Qual. score	1000g	dbsnp135
ADNSHL												
SB14-30	WFS1	mis.	NM_006005	4	exon8	c.T1235C	p.V412A	118	131	99	0.0037	rs144951440
SB50-94	COCH	mis.	NM_001135058	14	exon4	c.T341C	p.L114P	18	10	99	-	-
SB54-101	OTOR	ns.	NM_020157	20	exon2	c.G223T	p.E75X	47	49	99	-	-
SB60-107	MYO6	ns.	NM_004999	6	exon8	c.C613T	p.R205X	108	83	99	-	-
SB61-109	COL11A2	mis	NM_080680	6	exon30	c.C2336T	p.P779L	85	69	99	0.0005	rs150877886
SH14-37	COCH	mis.	NM_001135058	14	exon3	c.G113A	p.G38D	79	78	99	-	-
SH20-47	EYA4	mis.	NM_172103	6	exon11	c.C909G	p.F303L	69	52	99	-	-
SH21-50	MYO6	ns.	NM_004999	6	exon8	c.C613T	p.R205X	41	51	99	-	-
SH40-89	GJB3	mis.	NM_001005752	1	exon2	c.G250A	p.V84I	125	123	99	0.0018	rs145751680
ARNSHL												
SB04-11	OTOF	Fs. del.	NM_194322	2	exon24	c.3133delC	p.R1045Gfs*28	10	7	99	-	-
	OTOF	ns.	NM_194322	2	exon8	c.C1122G	p.Y374X	75	52	99	-	-
SB38-75	STRC	ns.	NM_153700	15	exon20	c.C4057T	p.Q1353X	0	39	81.2	-	rs2614824
SB47-91	MYO3A	mis.	NM_017433	10	exon7	c.C580A	p.P194T	117	109	99	-	-
	MYO3A	Fs. ins.	NM_017433	10	exon16	c.1582_1583 insT	p.Y530Lfs*9	22	13	99	-	-
SH35-75	GJB2	Fs. del.	NM_004004	13	exon2	c.299_300del	p.H100Rfs*14	110	86	99	-	rs111033204
	GJB2	Fs. del.	NM_004004	13	exon2	c.235delC	p.L79Cfs*3	123	104	99	0.0023	rs80338943

#Cov.: coverage, Qual.: quality, mis.:missense, ns.:nonsense, Fs.del.:frameshift deletion, Fs.ins.:frameshift insertion.

Protein domain and 3D structure of the candidate genes were further analyzed. p.G38D and p.L114P of COCH in proband SH14-37 and SB50-94 were located in LCCL domain which had been studied as substantial for DFNA9 (Figure 2-9 and 2-10). Truncated form of mutation, p.R205X of MYO6 in SB60-107 and SH21-50 was located in the mid of motor domain of myosin head (Figure 2-11). p.V412A of WFS1 in SB14-30 was located in one of the helical transmembrane regions. Truncated form of mutation, p.E75X of OTOR lied in SH3 domain. p.F303L of EYA4 in SH20-47 was located in about 50 amino acid distance from active site among 639 total amino acids, but there was no 3D structure studied. p.V84I of GJB3 in SH40-89 was located in helical transmembrane domain. Compound heterozygous mutation of OTOF in SB04-11 was nonsense mutation in one allele and frameshift deletion in the other allele. Homozygous mutation of STRC in SB38-75 was nonsense mutation at Q1353 from total 1775 amino acids.

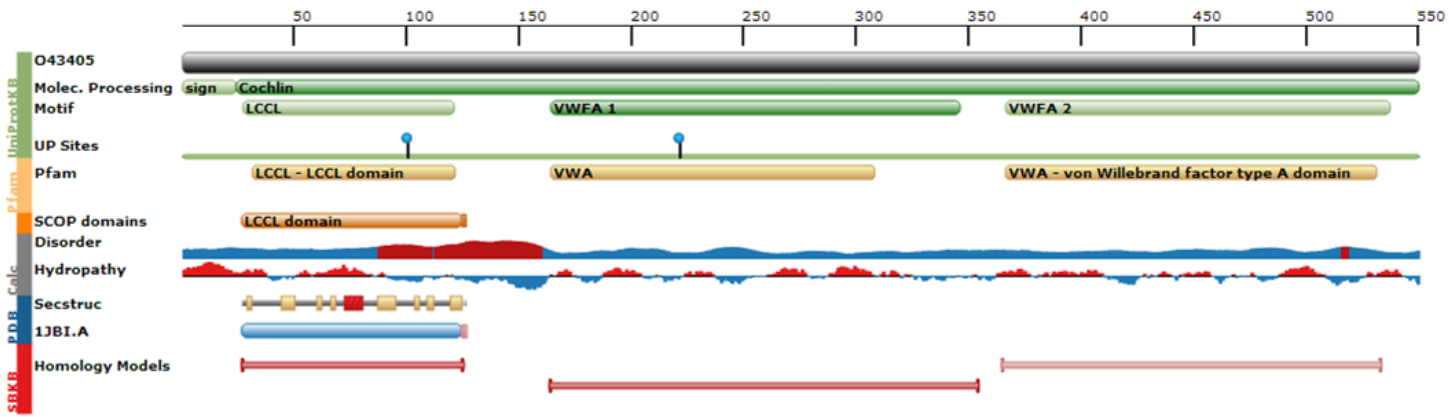


Figure 2-9. **Domains of human COCH including LCCL domain** (amino acid 27-124) that is substantial for DFNA9. p.G38D and p.L114P are located in LCCL domain.

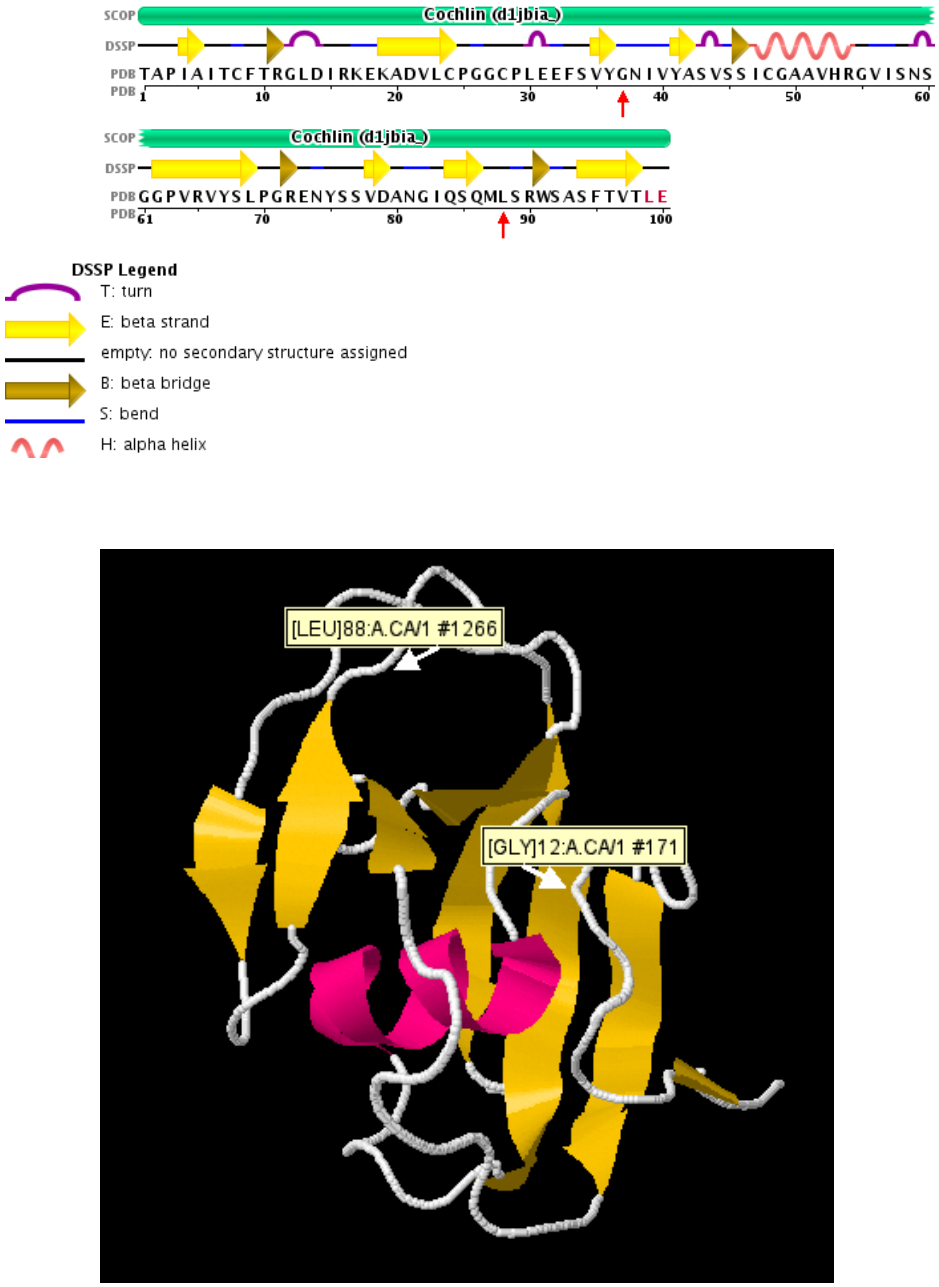


Figure 2-10. LCCL domain of COCH that has p.G38D and p.L114P mutation (red and white arrows).

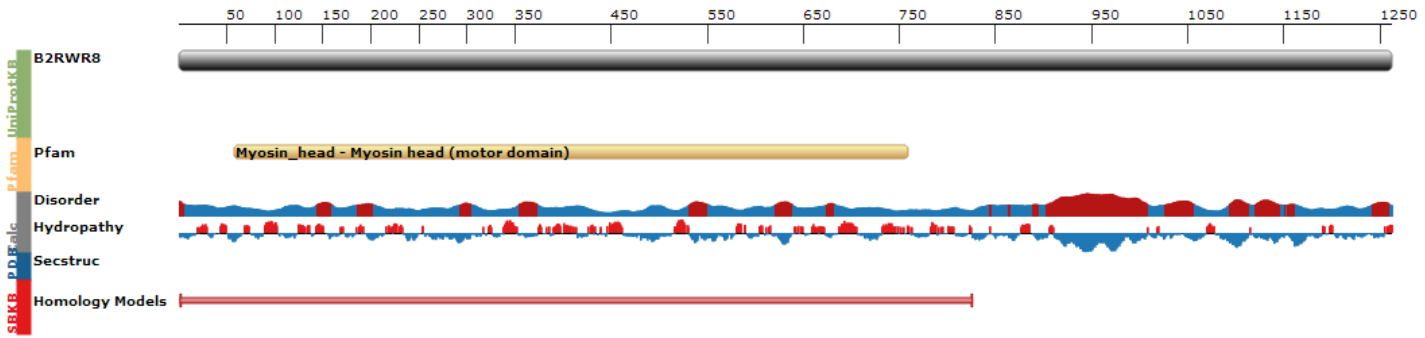


Figure 2-11. **Domains of myosin-6 of mouse including myosin head** (motor domain). p.R205X is located in motor

In summary, among 32 familial NSHL cases, I could detect mutations in 25 probands (79.1%) by Sanger and targeted exome sequencing in total. By the inheritance pattern, I was able to make a molecular genetic diagnosis from 9 (69.2%) of 13 autosomal dominant families on seven genes such as WFS1, COCH, EYA4, MYO6, GJB3, COL11A2 and OTOR, and from 9 (75.0%) of 12 autosomal recessive families on SLC26A4, GJB2, MYO3A, OTOF, and STRC. I also found one case with MRNR1 mutation with maternal inheritance, and five cases of POU3F4 mutation with X-linked inheritance. However, I could not detect candidate mutations in seven probands, in which the number of variants from basic filtering was not correlated with read depth in 20 probands (Figure 2-12). The number of called variants, sequencing depth and mean coverage was not different from those with candidate mutations detected (Figure 2-13).

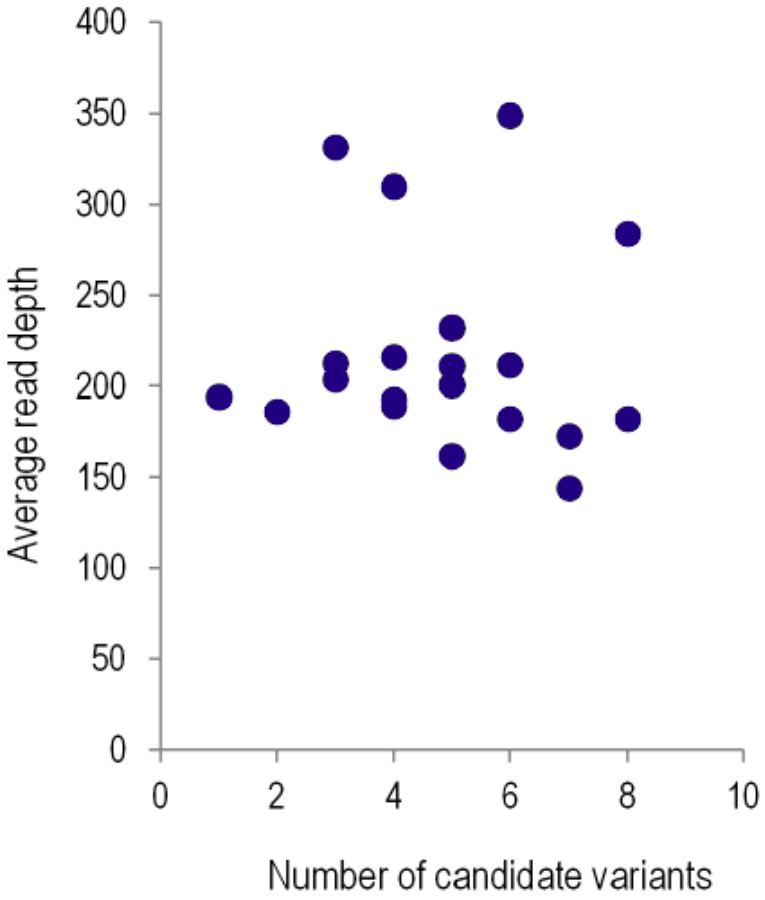


Figure 2-12. **Number of candidate SNVs and read depth**. The relationship between the numbers of candidate SNVs and read depth were plotted in 20 probands

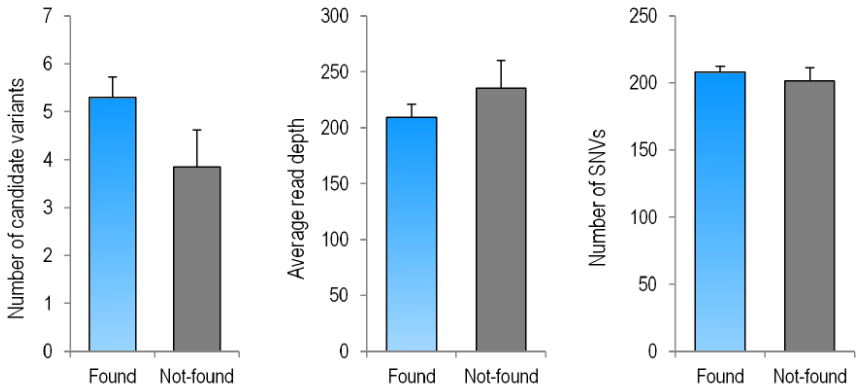


Figure 2-13. Statistics between the two groups (found and not found). Candidate SNV-found patient group (Found) was compared with patient group without candidate SNV (Not-found) in the number of candidate variants, read depth and called SNVs

DISCUSSION

Genetic cause of sensorineural disorders such as mental retardation, retinitis pigmentosa, and congenital hearing loss is extraordinarily heterogeneous. It is often hard to detect disease-causing mutation in each patient because I have to screen all the candidate genes. However, the genetic diagnosis by highthroughput sequence analysis gives clinicians and patients the opportunity to delineate the characteristics of disease. Especially early intervention of hearing loss in children might provide better clinical outcomes in the linguistic development. In this study, I could enhance the efficiency to find genetic alterations in familial NSHL patients. A candidate gene approach using conventional PCR sequencing against the candidate genes related to a certain phenotypic marker can cover only 10–20% of familial NSHL cases. Methodologies that enable us to effectively screen these common mutations related to the certain phenotypic markers, such as multiplex SNaPshot minisequencing, have been introduced in our population [7]. Howerver, a substantial portion of NSHL cases without any phenotypic marker still remains unanswered in terms of the molecular genetic etiology.

Therefore, I propose new diagnostic pipeline with high sensitivity to detect candidate mutations (Figure 2-1).

Through the targeted resequencing of the 20 families, I found most likely responsible genes for nine out of thirteen AD families and four genes from seven AR families. Nevertheless, seven cases still need to find the final candidate mutation. The probands SB41-78 and SB55-102 have a relatively

subtle phenotype considering their ages (35 years old and 59 years old, respectively). Therefore, it is possible that their hearing loss is just a phenocopy. The proband SB40-77 showed characteristic mid frequency hearing loss, which rendered us to focus upon the candidate autosomal dominant genes such as TECTA or COL11A2 that has been reported to cause mid frequency hearing loss. However, I could not detect a candidate variant in those genes. Rather, I found a potentially pathogenic variant (c.G5054A:p.R1685Q) in the MYH14 gene, a known deafness gene in DFNA4 locus from this proband. This variant has been detected neither in the 160 normal Korean control chromosomes nor in 1000 genomes. It was predicted to be 'probably damaging' by the Polyphen. In addition, the R1685 residue was conserved among many species including several mammalians, frog and zebrafish. However the audiogram pattern is not compatible with the previous reports [8,9] and I was not able to check the segregation of the variant due to other family members' reluctance to participate in this study. Mutations in the different domains in the MYH14 might lead to different audiogram configurations. Currently, I am thinking that this MYH14 variant might account for the phenotype but did not count this as a causative mutation for this study. As for the family SH41, p.A2T variant of the OTOR gene (Accession No. AF233261) was detected after the basic filtering. This variant was not detected neither of normal 160 Korean chromosomes nor 1000 genomes. Robertson et al. (2000) proposed OTOR's possible role in human deafness based upon its preferential and abundant expression in the cochlea [10]. However, this variant did not cosegregate with hearing loss in one of the

member in the family SH41.

Recently, narrow bony cochlear nerve canal (nBCNC) has been recognized and spotlighted as the most frequent inner ear anomaly [11]. Our group has postulated that the bilateral nBCNC may have a genetic etiology while the unilateral nBCNC is least likely to have a genetic contribution [12]. However, I was not able to find any candidate variant among the 80 deafness genes in the family SH27 where there was a sibling pair with bilateral nBCNC. The family SH23-52 segregates hearing loss presumably in an autosomal recessive manner, since the parents of three affected children showed perfect normal hearing. It is likely that the causative mutations for hearing loss in these families reside in genes other than 80 genes in this panel. I will further analyze the mutation in all exome of each family, because it may not be present in 80 candidate genes studied here.

Another reason for the detection failure may be due to the technical incompleteness. Coverage of targeted sequencing is not perfect to miss some exons, but usually considered good enough or not for the further analysis, especially in the experiments with many targets. This study also showed that 10% of exons were not properly captured. Capturing efficiency will be increased by new technologies for next generation sequencing. Recently, new enrichment technologies such as a semi-automated PCR amplification or a microdroplet PCR- based approach replacing the conventional hybridization-based enrichment technique have been successfully utilized in combination with next generation sequencing for genetic diagnosis of familial autosomal recessive deaf patients [13,14]. However, the diagnostic yield in these studies

was not greatly different from those in ours and previous studies utilizing the hybridization –based enrichment technique [15,16], rendering us to believe that technical incompleteness in capturing cannot solely account for the detection failure.

I found most likely responsible genes for nine out of thirteen AD families. Among seven autosomal recessive NSHL cases, I could detect the mutations in four genes such as GJB2, MYO3A, STRC and OTOF in four cases. One of them was mutations in GJB2, which were used as a positive control because it is well known for causing severe prelingual hearing loss as for the patient SH35-75. The sequence analysis of candidate genes may be easier to use PCR method, but clinical decision for candidate gene sequencing may be more difficult for all the clinician. To this end, I propose simple single step sequence analysis using targeted exome sequencing.

REFERENCES

General Introduction

1. Harlor AD Jr, Bower C; Committee on Practice and Ambulatory Medicine; Section on Otolaryngology-Head and Neck Surgery. Hearing assessment in infants and children: recommendations beyond neonatal screening. Pediatrics. 2009 Oct;124(4):1252-63.

2. 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(2–3):189–196.

3. Dror AA, Avraham KB: Hearing impairment: a panoply of genes and functions. Neuron 2010, 68(2):293–308.

4. Hilgert N, Smith RJ, Van Camp G (2009) Function and expression pattern of nonsyndromic deafness genes. Curr Mol Med 9: 546–564.

5. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013 Oct;14(10):681-91.

6. Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. Hear Res. 2012 Jun;288(1-2):67-76.

7. Schraders M, Haas SA, Weegerink NJ, Oostrik J, Hu H, Hoefsloot LH, Kannan S, Huygen PL, Pennings RJ, Admiraal RJ, Kalscheuer VM, Kunst HP,

Kremer H. Next-generation sequencing identifies mutations of SMPX, which encodes the small muscle protein, X-linked, as a cause of progressive hearing impairment. Am J Hum Genet. 2011 May 13;88(5):628-34.

8. De Keulenaer S, Hellemans J, Lefever S, Renard JP, De Schrijver J, Van de Voorde H, Tabatabaiefar MA, Van Nieuwerburgh F, Flamez D, Pattyn F, Scharlaken B, Deforce D, Bekaert S, Van Criekinge W, Vandesompele J, Van Camp G, Coucke P. Molecular diagnostics for congenital hearing loss including 15 deafness genes using a next generation sequencing platform. BMC Med Genomics. 2012 May 18;5:17.

9. Baek JI, Oh SK, Kim DB, Choi SY, Kim UK, Lee KY, Lee SH. Targeted massive parallel sequencing: the effective detection of novel causative mutations associated with hearing loss in small families. Orphanet J Rare Dis. 2012 Sep 3;7:60.

Chapter 1.

1. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011, 12(11):745–755.

2. Yang W, Ahmed M, Elian M, Hady E-SA, Levchenko TS, Sawant RR, Signoretti S, Collins M, Torchilin VP, Goldberg SN: Do Liposomal Apoptotic Enhancers Increase Tumor Coagulation and End-Point Survival in Percutaneous Radiofrequency Ablation of Tumors in a Rat Tumor Model? Radiology 2010, 257(3):685–696.

3. Min BJ, Kim N, Chung T, Kim OH, Nishimura G, Chung CY, Song HR,

Kim HW, Lee HR, Kim J, et al: Whole-exome sequencing identifies mutations of KIF22 in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. Am J Hum Genet 2011, 89(6):760–766.

4. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, et al: Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N Engl J Med 2010, 363(23):2220–2227.

5. Gilissen C, Hoischen A, Brunner HG, Veltman JA: Disease gene identification strategies for exome sequencing. Eur J Hum Genet 2012, 20(5):490–497.

6. Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, et al: Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. Nat Genet 2010, 42(10):827–829.

7. Cooper GM, Shendure J: Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet 2011, 12(9):628–640.

8. Nielsen R, Paul JS, Albrechtsen A, Song YS: Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 2011, 12(6):443–451.

9. Smith KR, Bromhead CJ, Hildebrand MS, Shearer AE, Lockhart PJ, Najmabadi H, Leventer RJ, McGillivray G, Amor DJ, Smith RJ, et al: Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biol 2011, 12(9):R85.

10. Cho TJ, Kim OH, Choi IH, Nishimura G, Superti-Furga A, Kim KS, Lee YJ, Park WY: A dominant mesomelic dysplasia associated with a 1.0-Mb microduplication of HOXD gene cluster at 2q31.1. J Med Genet 2010, 47(9):638–639.

11. Sha BY, Yang TL, Zhao LJ, Chen XD, Guo Y, Chen Y, Pan F, Zhang ZX, Dong SS, Xu XH, et al: Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. J Hum Genet 2009, 54(4):199–202.

12. Consortium IS: Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008, 455(7210):237–241.

13. Kang GC, Gan AW, Yam A, Tan AB, Tay SC: Mycobacterium abscessus Hand Infections in Immunocompetent Fish Handlers: Case Report. J Hand Surg Am 2010, 35(7):1142–1145.

14. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, Tothill RW, Halgamuge SK, Campbell IG, Gorringe KL: CONTRA: copy number analysis for targeted resequencing. Bioinformatics 2012, 28(10):1307–1313.

15. 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(2–3):189–196.

16. Dror AA, Avraham KB: Hearing impairment: a panoply of genes and functions. Neuron 2010, 68(2):293–308.

17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: A method and server for predicting

damaging missense mutations. Nat Methods 2010, 7(4):248–249.

18. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010, 7(8):575–576.

19. Hollox EJ, Barber JC, Brookes AJ, Armour JA: Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23.1. Genome Res 2008, 18(11):1686–1697.

20. Carelle-Calmels N, Saugier-Veber P, Girard-Lemaire F, Rudolf G, Doray B, Guerin E, Kuhn P, Arrive M, Gilch C, Schmitt E, et al: Genetic compensation in a human genomic disorder. N Engl J Med 2009, 360(12):1211–1216.

21. Huang RS, Chen P, Wisel S, Duan S, Zhang W, Cook EH, Das S, Cox NJ, Dolan ME: Population-specific GSTM1 copy number variation. Hum Mol Genet 2009, 18(2):366–372.

22. Cherny SS, Abecasis GR, Cookson WO, Sham PC, Cardon LR: The effect of genotype and pedigree error on linkage analysis: analysis of three asthma genome scans. Genet Epidemiol 2001, 21(Suppl 1):S117–S122.

23. Sonnemann KJ, Fitzsimons DP, Patel JR, Liu Y, Schneider MF, Moss RL, Ervasti JM: Cytoplasmic gamma-actin is not required for skeletal muscle development but its absence leads to a progressive myopathy. Dev Cell 2006, 11(3):387–397.

24. Drummond MC, Belyantseva IA, Friderici KH, Friedman TB: Actin in hair cells and hearing loss. Hear Res 2012, 288(1–2):89–99.

25. Riviere JB, van Bon BW, Hoischen A, Kholmanskikh SS, O'Roak BJ,

Gilissen C, Gijsen S, Sullivan CT, Christian SL, Abdul-Rahman OA, et al: De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. Nat Genet 2012, 44(4):440–444. S441-442.

26. Morell RJ, Friderici KH, Wei S, Elfenbein JL, Friedman TB, Fisher RA: A new locus for late-onset, progressive, hereditary hearing loss DFNA20 maps to 17q25. Genomics 2000, 63(1):1–6.

27. Zhu M, Yang T, Wei S, DeWan AT, Morell RJ, Elfenbein JL, Fisher RA, Leal SM, Smith RJ, Friderici KH: Mutations in the gamma-actin gene (ACTG1) are associated with dominant progressive deafness (DFNA20/26). Am J Hum Genet 2003, 73(5):1082–1091.

28. van Wijk E, Krieger E, Kemperman MH, De Leenheer EM, Huygen PL, Cremers CW, Cremers FP, Kremer H: A mutation in the gamma actin 1 (ACTG1) gene causes autosomal dominant hearing loss (DFNA20/26). J Med Genet 2003, 40(12):879–884.

29. Belyantseva IA, Perrin BJ, Sonnemann KJ, Zhu M, Stepanyan R, McGee J, Frolenkov GI, Walsh EJ, Friderici KH, Friedman TB, et al: Gamma-actin is required for cytoskeletal maintenance but not development. Proc Natl Acad Sci U S A 2009, 106(24):9703–9708.

30. Perrin BJ, Sonnemann KJ, Ervasti JM, Perrin BJ, Sonnemann KJ, Ervasti JM: beta-actin and gamma-actin are each dispensable for auditory hair cell development but required for Stereocilia maintenance. PLoS Genet 2010, 6(10):e1001158.

31. Rendtorff ND, Zhu M, Fagerheim T, Antal TL, Jones M, Teslovich TM, Gillanders EM, Barmada M, Teig E, Trent JM, et al: A novel missense

mutation in ACTG1 causes dominant deafness in a Norwegian DFNA20/26 family, but ACTG1 mutations are not frequent among families with hereditary hearing impairment. Eur J Hum Genet 2006, 14(10):1097–1105.

32. Baek JI, Oh SK, Kim DB, Choi SY, Kim UK, Lee KY, Lee SH: Targeted massive parallel sequencing: the effective detection of novel causative mutations associated with hearing loss in small families. Orphanet J Rare Dis 2012, 7(1):60.

33. de Heer AM, Huygen PL, Collin RW, Oostrik J, Kremer H, Cremers CW: Audiometric and vestibular features in a second Dutch DFNA20/26 family with a novel mutation in ACTG1. Ann Otol Rhinol Laryngol 2009, 118(5):382–390.

34. Liu P, Li H, Ren X, Mao H, Zhu Q, Zhu Z, Yang R, Yuan W, Liu J, Wang Q, et al: Novel ACTG1 mutation causing autosomal dominant non-syndromic hearing impairment in a Chinese family. J Genet Genomics 2008, 35(9):553–558.

35. Khaitlina SY: Functional specificity of actin isoforms. Int Rev Cytol 2001, 202:35–98.

36. Cho TJ, Kim OH, Lee HR, Shin SJ, Yoo WJ, Park WY, Park SS, Cho SI, Choi IH: Autosomal recessive multiple epiphyseal dysplasia in a Korean girl caused by novel compound heterozygous mutations in the DTDST (SLC26A2) gene. J Korean Med Sci 2010, 25(7):1105–1108.

37. Vigeland MD: paramlink: Parametric linkage analysis in R. In R package version 06–1. 2012. http://CRAN.R-project.org/package=paramlink.

38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H,

Shindyalov IN, Bourne PE: The Protein Data Bank. Nucleic Acids Res 2000, 28(1):235–242.

39. Otterbein LR, Graceffa P, Dominguez R: The crystal structure of uncomplexed actin in the ADP state. Science 2001, 293(5530):708–711.

40. Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, Wagener J, Murray-Rust P, Steinbeck C, Wikberg JE: Bioclipse: an open source workbench for chemo- and bioinformatics. BMC Bioinforma 2007, 8:59.

41. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013, 14(10):681-91.

Chapter 2.

1. Hilgert N, Smith RJ, Van Camp G (2009) Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? Mutat Res 681: 189–196.

2. Resendes BL, Williamson RE, Morton CC (2001) At the speed of sound: gene discovery in the auditory system. Am J Hum Genet 69: 923–935.

3. Dror AA, Avraham KB (2010) Hearing impairment: a panoply of genes and functions. Neuron 68: 293–308.

4. Hilgert N, Smith RJ, Van Camp G (2009) Function and expression pattern of nonsyndromic deafness genes. Curr Mol Med 9: 546–564.

5. Lin X, Tang W, Ahmad S, Lu J, Colby CC, et al. (2012) Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. Hear Res 288: 67–76.

6. Min BJ, Kim N, Chung T, Kim OH, Nishimura G, et al. (2011) Whole-exome sequencing identifies mutations of KIF22 in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. Am J Hum Genet 89: 760–766.

7. Sagong B, Baek JI, Oh SK, Na KJ, Bae JW, et al. (2013) A rapid method for simultaneous screening of multi-gene mutations associated with hearing loss in the Korean population. PLoS One 8: e57237.

8. Donaudy F, Snoeckx R, Pfister M, Zenner HP, Blin N, et al. (2004) Nonmuscle myosin heavy-chain gene MYH14 is expressed in cochlea and mutated in patients affected by autosomal dominant hearing impairment (DFNA4). Am J Hum Genet 74: 770–776.

9. Yang TH, Young YH (2005) Click-evoked myogenic potentials recorded on alert guinea pigs. Hear Res 205: 277–283.

10. Robertson NG, Heller S, Lin JS, Resendes BL, Weremowicz S, et al. (2000) A novel conserved cochlear gene, OTOR: identification, expression analysis, and chromosomal mapping. Genomics 66: 242–248.

11. Adunka OF, Jewells V, Buchman CA (2007) Value of computed tomography in the evaluation of children with cochlear nerve deficiency. Otol Neurotol 28: 597–604.

12. Oh SK, Choi SY, Yu SH, Lee KY, Hong JH, et al. (2013) Evaluation of the pathogenicity of GJB3 and GJB6 variants associated with nonsyndromic hearing loss. Biochim Biophys Acta 1832: 285–291.

13. Schrauwen I, Sommen M, Corneveaux JJ, Reiman RA, Hackett NJ, et al. (2013) A sensitive and specific diagnostic test for hearing loss using a microdroplet PCR-based approach and next generation sequencing. Am J Med

Genet A 161A: 145–152.

14. De Keulenaer S, Hellemans J, Lefever S, Renard JP, De Schrijver J, et al. (2012) Molecular diagnostics for congenital hearing loss including 15 deafness genes using a next generation sequencing platform. BMC Med Genomics 5: 17.

15. Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, et al. (2011) Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families. Genome Biol 12: R89.

16. Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, et al. (2010) Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc Natl Acad Sci U S A 107: 21104–21109.

국문 초록

서론: 난청은 비교적 흔한 신경감각기관의 질환이며, 절반 정도가 유전적 원인이 이유가 된다. 청각기전이 복잡하므로 이전 연구들의 결과로는 50개 정도의 유전자가 결함이 있는 것으로 보고되고 있다. 차세대염기서열분석은 DNA의 대량 병렬서열분석을 통한 기술로 큰 데이터로부터 유전정보와 변이를 밝히는 데 이용이 된다. 차세대염기서열분석의 몇 가지 분석법으로부터 난청의 원인 변이를 좁혀 밝히는 것이 가능하다.

방법: 두 가지 독립적인 접근법이 가족성 비증후군 난청의 다른 환자 그룹에 대해서 사용되었다. 하나의 대가족에 대해 전장엑솜서열분석이, 그리고 다수의 난청환자들에 대해 타겟리시퀀싱이 각각 사용되었다. 전장엑솜서열분석은 4 명의 환자와 4 명의 정상가족에 대해 이루어졌고, 유전자복제수변이, 연관분석, 단일염기변이에 대한 다층분석이 행해졌다. 타겟리시퀀싱은 이전의 스크리닝을 거쳐 선정된 20 명의 환자에 대해 이루어졌고 원인변이를 찾기위해 다단계 조건을 두어 후보변이를 걸렀다.

결과: 전장엑솜서열분석에서, 질환과 같이 분리되는 후보 유전자복제수변이는 없었다. 연관분석에서 LOD score > 1 인 6 개의 지역을 발견하였고, 감마액틴 1 의 한 단일염기변이가 새롭게 발견된 돌연변이였다. 감마액틴 1 은 난청의 알려진 원인 유전자이고 돌연변이된

지역은 ATP 와 상호작용하는 것으로 예측된 것이었다. 타겟리시퀀싱에서, 전체 대상은 원래 32 명의 환자였다. 스크리닝 유전분석에 실패한 20 명 환자들이 타겟유전자 내의 단일염기변이와 삽입, 결실에 대해 분석되었다. 유전자의 유전되는 양상, 환자와 정상대조군과 가족에 대한 Sanger 시퀀싱, 그리고 임상양상을 가지고 다단계로 후보변이들을 걸렀다. 20 명 중 13 명의 환자들이 최종 후보변이를 가졌고 결과적으로 총 32 명 중 25 명이 후보변이를 가졌다.

결론: 난청의 정확한 유전진단이 가능하다면, 잠정적인 치료의 시도나 유전상담이 고려될 수 있다. 특히 차세대염기서열분석의 분석법으로 난청의 유전진단이 좀더 접근이 쉽게 되고 효율적이 될 수 있을 것이다.

* 본 내용은 "Multiphasic analysis of whole exome sequencing data identifies a novel mutation of ACTG1 in a nonsyndromic hearing loss family. BMC Genomics. 2013 Mar 18;14:191." 과 "Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss. PLoS One. 2013 Aug 22;8(8):e68692." 에 출판 완료된 내용임

주요어: 난청, 차세대염기서열분석, 전장엑솜서열분석, 타겟리시퀀싱
학번: 2010-21905