Comprehensive Genetic Profiling of Sensorineural Hearing Loss Using an Integrative Diagnostic Approach

Sang-Yeon Lee[1,2,3,8], Seungbok Lee[2,4,8], Seongyeol Park[5,8], Sung Ho Jung[1], Yejin Yun[1], Won Hoon Choi[1], Ju Hyuen Cha[1], Hongseok Yun[2], Sangmoon Lee[5], Myung-Whan Suh[1], Moo Kyun Park[1], Jae-Jin Song[6], Byung Yoon Choi[6], Jun Ho Lee[1], Young Seok Ju[5,7], June-Young Koh[5,*], Jong-Hee Chae[2,4,*]

[1] Department of Otorhinolaryngology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, South Korea

[2] Department of Genomic Medicine, Seoul National University Hospital, Seoul, South Korea

[3] Sensory Organ Research Institute, Seoul National University Medical Research Center

[4] Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, South Korea

[5] Inocras Inc., San Diego, California, United States

[6] Department of Otorhinolaryngology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, South Korea

[7] Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

[8] These authors contributed equally to this work.

*** Corresponding author**

June-Young Koh

Inocras, Inc., 6330 Nancy Ridge Drive Suite 106, San Diego, CA 92121, United States

Tel : +1-858-665-2120

E-mail: jy.koh@inocras.com

Jong-Hee Chae

Division of Pediatric Neurology, Department of Pediatrics,

Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital,

Seoul National University College of Medicine

101 Daehakro Jongno-gu, Seoul, South Korea 03080

Tel : +82-2-2072-3622; Fax : +82-2-7433455

E-mail: chaeped1@snu.ac.kr

Total number of words

Abstract (188), Main Text (3,756)

Abstract

Despite the advent of Next-Generation Sequencing (NGS), genetic diagnosis of genetic disorders remains challenging, with diagnostic rates plateauing at approximately 50%. We investigated sensorineural hearing loss (SNHL), a prevalent sensory disorder with substantial genetic heterogeneity, through a comprehensive genomic analysis of a homogeneous disease cohort. Leveraging 394 families (750 individuals), we implemented a systematic multi-tiered genomic approach encompassing single-gene analysis to whole-genome sequencing (WGS), integrated with functional assays and bioinformatic analysis. Our methodological framework revealed a cumulative diagnostic yield of 55.6% (219 families), with automated WGS bioinformatics pipeline uncovering an additional 20 families harboring pathogenic variants, predominantly structural variants. Notably, comparative genomic analysis unveiled a higher frequency of single pathogenic alleles in recessive genes within our SNHL cohort relative to control populations. Subsequent deep intronic region interrogation identified three pathogenic variants on the opposite allele, substantiating the diagnostic utility of comprehensive genomic profiling. Through this approach, we delineated a genome-phenome landscape of SNHL, elucidating molecular signatures and establishing genotype-phenotype correlations at the inner ear functional level. This study underscores the transformative potential of WGS as a robust molecular diagnostic modality, advancing precision medicine paradigms in genetic disease research.

Keywords: Sensorineural hearing loss; Whole-genome sequencing; Stepwise genomic approach; Genotype-phenotype correlations

Introduction

Hearing is the primary sense used for human communication and an important component in the development of language and music. Thus, hearing impairment, the most common sensory deficit in humans, is a major public health problem, affecting approximately 466 million people worldwide (World Health Organization, https://who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss). Sensorineural hearing loss (SNHL)—i.e., defective sound signaling in the auditory sensory system—can be caused by multiple etiologies, including genetic causes, congenital infections, trauma, ototoxic medications, and autoimmune disorders[1]. Since the 2010s, advances in high-throughput next-generation sequencing (NGS) technologies have facilitated extensive elucidation of the genetic backgrounds of SNHL, with a focus on monogenic forms of deafness. Notably, mouse genetics studies have helped reveal the physiological basis of SNHL in humans and the associated molecular functions[2].

Despite growing recognition of the significance of genetic diagnosis of SNHL, it remains challenging to identify a genetic diagnosis in SNHL with substantial genetic heterogeneity[3,4]. NGS is increasingly favored for genetic diagnosis due to its capacity for simultaneous large-scale genetic loci screening, and methods like targeted panel sequencing (TPS) and whole-exome sequencing (WES) were widely used in real-world practice[5-7]. In the literature, targeted sequencing for SNHL has achieved diagnostic yields of between 12.7% and 64.3%[8-11]. However, even after exome sequencing, approximately 50% of cases remain genetically elusive.

As the cost of sequencing dramatically declines[12], the clinical application of whole-genome sequencing (WGS) becomes more feasible, which has a higher capability to detect a more diverse spectrum of genomic variants that had not

previously been captured by exome sequencing or other targeted approches[13,14]. Recent studies have shown the clinical utility of WGS for the genetic diagnosis of several disorders and effectively shortening their diagnostic odyssey, increasingly considered as a first-line genetic test[15-19]. However, WGS is not yet widely applied in routine clinical settings for diagnosing patients with rare diseases, including SNHL, due to several limitations, such as the difficulties of rapid bioinformatic analysis and accurate clinical interpretation. Additionally, although whole genomes are sequenced, the analysis is often limited to in silico gene panels or the coding regions of the genome[20].

In the present study, we comprehensively explored the genetic landscape of 394 prospective SNHL families. Using a stepwise approach from single target gene analysis to WGS, we evaluated the additional diagnostic value of WGS. We implemented an automated WGS bioinformatics pipeline, integrating *in-house* algorithms with manual curation by both otologists and medical geneticists. This approach allowed comprehensive analysis of all variant types, and improved the diagnostic yield for previously undiagnosed patients. Further analysis of deep intronic regions identified novel pathogenic variants. These findings refined the genotype-phenotype landscape of SNHL, revealing the gene signatures based on phenotypes and identifying genotype-phenotype correlations at the level of inner ear molecular functions. Our results demonstrate the clinical utility of an integrated molecular diagnostic approach, including WGS, in real-world SNHL practice, paving the way toward precision medicine.

Results

A stepwise approach of genetic testing for patients with SNHL

We conducted stepwise approach of genetic testing, sanger sequencing, TPS, WES, mitochondrial DNA (mtDNA) sequencing, multiplex ligation-dependent probe amplification (MLPA), and WGS, to prospectively recruited SNHL cohort (*n* of probands = 394; n of participants including probands and their family members = 750; **Supplementary Table 1**), including non-syndrome hearing loss (ns-SNHL; *n* = 341, 86.5%), and syndromic hearing loss (s-SNHL; n = 53, 13.5%). Genetic testing was structured into five sequential steps (**Fig. 1 and Supplementary Fig. 1a**). In Step 1, we screened 22 variants from 10 classical deafness genes (*GJB2, SLC26A4*, *TMPRSS3*, *CDH23*, *OTOF*, *TMC1*, *ATP1A3*, *MPZL2*, *COCH*, and 12S rRNA). Among cohort, ns-SNHL patients (*n* = 341) were initially subjected to Step 1 and the ns-SNHL patients who remained undiagnosed after Step 1 (*n* = 295), and s-SNHL patients (*n* = 53) were then subjected to the next steps (Steps 2-1 and 2-2). In Step 2-1, 99 and 249 patients with SNHL were subjected to TPS (including 246 genes) and WES, respectively. After that, undiagnosed patients with suggestive clinical features, including symmetric mild-to-moderate SNHL (*n* = 51)[21], apparent branchio-oto-renal/branchio-otic (BOR/BO) syndrome (*n* = 2)[22], radiological evidence of enlarged vestibular aqueduct (EVA) (*n* = 1), and suspicious mitochondria disorders (*n* = 11) were subjected to Step 2-2, including MLPA methods (*n* = 54) or mtDNA panel sequencing (*n* = 11). In Step 3-1, among patients who remained undiagnosed, all s-SNHL patients and a representative subset of ns-SNHL patients were selected for WGS using a well-thought-out sample size estimation with a stratified sampling approach (**Supplementary Fig. 1b**). For the remaining undiagnosed patients (*n* = 100), deep intronic regions of SNHL-related genes were screened to identify

additional pathogenic variants in Step 3-2. After stepwise genetic testing, patients with suspected pathogenic genetic causes underwent subsequent bioinformatic analyses and curation. Multidisciplinary molecular board meetings, comprising both clinicians and genome scientists, were then conducted to confirm genetic diagnosis.

Incremental improvement in diagnostic yield through stepwise genetic testing for SNHL

Through comprehensive genetic testing, we incrementally improved the diagnostic yield in the cohort, identifying disease-causing variants in 219 families (55.6%; **Fig. 2a**). Following the stepwise approach, in Step 1, we identified causal variants in 46 out of 341 patients (diagnostic yield for Step 1: 13,5%; cumulative yield: 11.7%). In Step 2-1, causal variants were found in 124 out of 348 patients (diagnostic yield for Step 2-1: 35.6%; cumulative yield: 43.1%), while Step 2-2 revealed causal variants in 26 out of 65 patients (diagnostic yield for Step 2-2: 40%; cumulative yield: 49.7%). WGS was performed on 120 patients, identifying causal variants in additional 23 probands (20 in Step 3-1 and 3 in Step 3-2; diagnostic yield for Step 3: 19.2%; cumulative yield: 55.6%). Details of individual patients, including the performed tests, diagnostic outcomes, and identified variants, are summarized in **Supplementary Table 2**.

Multivariate analysis revealed that the rate of genetic diagnosis was higher among patients with early identification of SNHL (adjusted odds ratio [OR], 1.32; 95% confidence interval [CI], 1.11–1.57) and those with a family history (adjusted OR, 1.55; 95% CI, 1.31–1.82; **Fig. 2b and Supplementary Table 3**). Similarly, patients with syndromic features were more likely to have identified genetic variants (adjusted OR, 1.44; 95% CI, 1.17–1.70). In contrast, a lower rate of genetic diagnosis was

observed in patients with adult-onset SNHL (adjusted OR, 0.50; 95% CI, 0.35–0.68), asymmetric hearing loss (adjusted OR, 0.38; 95% CI, 0.18–0.64), and interaural asymmetry (adjusted OR, 0.43; 95% CI, 0.23–0.68). Additionally, in the context of WGS, several features were associated with a higher likelihood of achieving a genetic diagnosis (**Fig. 2c and Supplementary Table 4**). The multivariate analyses suggested that genetic diagnosis was significantly related to the presence of syndromic features (adjusted OR, 2.51; 95% CI, 1.15–5.04) and early identification through a failed newborn hearing screening (NHS) (adjusted OR, 2.35; 95% CI, 1.13–4.97). In addition, the diagnostic yield significantly varied depending on the WGS approach. Trio-based WGS had a higher likelihood of identifying causal variants (adjusted OR, 3.71; 95% CI, 1.67–9.68), whereas singleton WGS was less effective (adjusted OR, 0.32; 95% CI, 0.12–0.71).

Further, we conducted the correlation analysis between the frequencies of causative variants among our cohort and population allele frequencies of the variants from public databases, such as the gnomAD[23] and KOVA[24] (Korean Variant Archive: Korean population database; **Fig. 2d**). We observed the strongest association with East Asians compared to other ethnicity groups within gnomAD, and a notably higher association with KOVA, suggesting that patient ethnicity should be considered in variant discovery. We discovered three outlier variants (*SLC12A3* p.Thr180Lys, *ESRRB* p.Arg382Cys, and *GJB2* p.Val37Ile) when comparing the variant frequencies with allele frequencies in KOVA (**Fig. 2e**). These variants have higher frequencies in the Korean population than in diagnosed patients, suggesting either their low penetrance features or the potential for functional pathogenicity despite its particularly higher allele frequencies[25].

In conclusion, from Step 1 to Step 3-2, we identified the genetic causes of

SNHL in 219 out of 394 probands (55.6%), with 23 diagnoses made through WGS. Among the identified variants, deep intronic variants (*n* = 3) and structural variants (SVs) (*n* = 11) were predominantly detected through WGS, and 13 variants of uncertain significance (VUS) required functional assays to validate their pathogenicity (**Fig. 2f**).

Comprehensive characterization of causative variants in SNHL

Using the genetic findings obtained through comprehensive stepwise genetic tests, we illustrated a mutational landscape of SNHL. Collectively, 63 genes were identified as disease-causing within 219 genetically diagnosed families (**Fig. 3a and Supplementary Fig. 2a**). *GJB2* was the most frequently affected gene (10.5%, 23/219), followed by five genes (*SLC26A4*, *STRC*, *USH2A*, *CDH23*, *and MPZL2)* that were found in at least ten unrelated families (collectively >40% of all diagnosed cases). Conversely, 29 SNHL-associated genes were detected from only one family (collectively 13.2%; 29/219; **Supplementary Fig. 2a**), suggesting that many more rare genes can cause SNHL. The inheritance patterns of the 63 genes included autosomal recessive (26/63 genes; affecting 132 families; found double hits, including homozygote and compound heterozygote variants), autosomal dominant (33/63 genes; affecting 73 families; found single hit), X-linked (4/63 genes; affecting 6 families), and mitochondrial (3/63 genes; affecting 8 families) (**Fig. 3a).** Within our cohort, three identified genes—*TECTA*, *COL4A3*, and *WFS1*—exhibited both autosomal recessive and autosomal dominant inheritance traits. Moreover, one patient exhibited dual genetic etiologies, inherited from their parents, harboring compound heterozygous variants (c.299del:p.His100LeufsTer12 and c.235del:p.Leu79CysfsTer3) in *GJB2* and a heterozygous variant

(c.113G>A;p.Gly38Asp) in *COCH*.

Fig. 3b and **Supplementary Fig. 2b** display the mutational landscapes of the 352 causative variants, including the variant type. Among the 352 identified variants, we found missense variants (157, 44.6%), nonsense variants (51, 14.5%), frameshift variants (63, 17.9%), and inframe variants (10, 2.8%) within the exome region. Additionally, splicing variants (27, 7.7%), SVs (36, 10.2%), and mitochondrial variants (8, 2.3%) were identified. Interestingly, we observed differences in the distribution of variant types across genes. In particular, consistent with previous reports, SVs accounted for (20/30, 66.7%) of the total variants in the *STRC* gene.

Among the 352 causative variants (190 when the same variants were collapsed), 83 variants were novel (**Fig. 3c**). All of the 352 variants were either pathogenic (P) or likely pathogenic (LP) according to the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) guidelines (**Supplementary Table 5**)[26]. A variety of functional studies—ranging from molecular modeling to minigene splicing assay—were performed to test 16 highly suggestive variants (**Supplementary Table 6**), leading to the reclassification of 13 variants from “uncertain significance” to “likely pathogenic”. An analysis of the variant type distribution among the 83 novel variants revealed that SVs, detected mostly through WGS, accounted for a much larger proportion (12.0%) compared to previously reported variants (2.8%) **(Fig. 3d)**. In addition, we observed an increasing trend in the proportion of patients with novel variants as the steps progressed from Step 1 to Step 3-2 (**Fig. 3e**).

Novel SVs identified from WGS include exonic deletions on *SPATA5* and *CLCNKA-CLCNKB* (**Figs. 3f and 3g**). In one patient (SNUH 799), WGS identified a small SV (c.2227-3015_2354+1415del) exclusively involving exon 15 of *SPATA5* (**Fig.**

3f), along with an *in trans* short frameshift deletion in exon 16 (c.2679_2680del). Notably, consistent with a previous case report[27], this patient exhibited systemic clinical manifestations, including bilateral moderately severe SNHL, intractable epilepsy with diffuse brain atrophy, and global developmental delay (**Supplementary Fig. 3a**). Next, we investigated the molecular consequences of the small SV (c.2227-3015_2354+1415del) on *SPATA5*-dependent bioenergetics using Seahorse assays (**Supplementary Fig. 3b**). The oxygen consumption rate (OCR) showed reduced respiratory function in patient fibroblasts, with significant decreases observed in basal respiration ($P < 0.001$), maximal respiration ($P < 0.001$), and ATP production ($P < 0.001$) compared to control mother fibroblasts (**Supplementary Fig. 3c**). These findings suggest that the small deletion in *SPATA5* (c.2227-3015_2354+1415del) contributes to impaired mitochondrial function leading to SNHL.

Advanced WGS approach revealed additional deep intronic variants

We sought to analyze the variant status of undiagnosed SNHL patients, even after the automated WGS pipeline (Step 3-1 in **Fig.1**) had been performed. Specifically, we examined the frequency of carriers for autosomal recessive SNHL-related genes (**Supplementary Fig. 4a**) in these undiagnosed patients and compared it to the carrier status in control cohorts (HC1: patients with hepatocellular carcinoma, n = 553, HC2: patients with breast cancer, n = 571; **Fig. 4a**). Throughout the analysis, we focused on identifying candidate pathogenic variants in accordance with ACMG guidelines (P/LP variants) and rare variants (minor allele frequency; MAF < 1%) predicted to have a high impact. These included transcript ablation, splice variants, start lost, stop gained/lost, frameshift variants, transcript amplification, feature elongation/truncation, and exon-disrupting SVs or transposable elements

(TEs). Subsequently, for SNHL patients with a pathogenic variant on a single allele after Step 3-1, we concentrated on identifying additional pathogenic variants on the opposite allele in Step 3-2. Specifically, *in silico* splicing variant predictions were performed using SpliceAI[28] on intronic variants (MAF < 1%) identified in SNHL carriers. This approach aimed to detect potential splice-affecting variants that may have been overlooked in previous analyses.

The analysis revealed that the carrier rate for pathogenic variants in SNHL-related genes was significantly higher in the undiagnosed SNHL patient group compared to the two control cohorts (**Fig. 4b**). Additionally, an evaluation of the MAF of the identified candidate pathogenic variants showed that variants found in the undiagnosed SNHL group were rarer than those in the control group (**Fig. 4c**). Subsequent screening for deep intronic variants in undiagnosed SNHL patients identified three *in trans* variants with SpliceAI prediction scores greater than 0.2. These included deep intronic variants in *USH2A* (c.7120+1475A>G, c.14134-3169A>G, and c.4628-26037A>G), each found in different patients (SNUH 485, SNUH 503, and SNUH 513, respectively; **Fig. 4d and Supplementary Fig. 4b-c**). To evaluate their pathogenicity, minigene assays were designed with specific splice donor (SD) and splice acceptor (SA) sites. The assays revealed that these variants induce aberrant splicing that leads to the inclusion of pseudoexons of varying sizes: 267 bp for c.7120+1475A>G, 50 bp for c.14134-3169A>G, and 96 bp for c.4628-26037A>G (**Fig. 4e-f**). The resulting aberrant transcripts are predicted to contain premature stop codons, leading to truncated, non-functional usherin protein (**Fig. 4g**). Combined with first hits of the coding variant in this *USH2A* gene (c.10712C>T, c.14835del, and coding deletion, respectively), these alleles of *USH2A* were inactivated in these patients.

285

286 ***Genotype-phenotype correlations through molecular function-based gene***

287 ***clustering***

288 Based on the genetic diagnoses identified through comprehensive genetic

289 analysis, we investigate the relationship between the affected genes and clinical

290 manifestations. We found that causal genes of SNHL were closely linked to specific

291 clinical manifestations, with 22 genes contributing to clinical manifestations in ≥3

292 families (**Supplementary Fig. 5a**). The genotype-phenotype map revealed a list of

293 SNHL genes that represent the major attributes of SNHL phenotypes, with over 50%

294 of affected patients harboring variants in the same gene (**Supplementary Fig. 5b**).

295 Although variants within the same causative gene can manifest allelic and clinical

296 heterogeneity[29,30], information about signature genes associated with distinct clinical

297 phenotypes may aid in conducting in-depth genetic analyses within the highly

298 heterogenous genetic landscape of SNHL.

299 Based on these genotype-phenotype correlation, we further analyzed the

300 phenotypic heterogeneity of SNHL according to causal genes considering the

301 functions of the genes in the inner ear. The 63 SNHL-related genes were classified

302 into eight categories based on inner ear molecular functions (**Fig. 5a and**

303 **Supplementary Fig. 5c**)[31]: (1) hair bundle development and function; (2) synaptic

304 transmission; (3) hair cell adhesion and maintenance; (4) cochlear ion homeostasis;

305 (5) transmembrane and extracellular matrix; (6) oxidative stress, autoinflammation,

306 and mitochondrial defect; and (7) transcriptional regulation. To evaluate the

307 relationships between genes across categories, we compared gene expression

308 patterns using publicly available transcriptome data from human inner ear organoids

309 and human cochlear and vestibular organs[32]. Comparative analysis of gene

expression patterns through perturbation testing revealed trends (*P* = 0.07) in expression across gene categories (**Fig. 5b**).

Interestingly, post-hoc analysis revealed that four phenotypic attributes (syndromic features, mixed hearing loss, and hearing loss onset) were significantly associated with pathogenic variants in the categories **(Figs. 5c-d and Supplementary Table 7)**. Causal genes more prevalent among patients with syndromic features (i.e., *EYA1*, *SIX1*, and *MT-TL1*) were significantly enriched in categories 6 and 7. These findings were in line with the cell type specificities of the genes, since genes in categories 6 and 7 are more broadly expressed in multiple organs and cell types (**Supplementary Fig. 6a**). Genes associated with mixed type manifestation (i.e., *POU3F4*, *EYA1*, and *SIX1*) were more prevalent in category 7 compared to in category 1 (*P* = 0.001). We speculate that the genes associated with transcriptional regulation may affect the development of the second branchial arch of the middle ear, and the third window of the inner ear, accounting for conductive components of hearing loss. Furthermore, genes associated with early identification of hearing loss onset were more predominant in category 4 compared to in category 3 (*P* =0.001). It is likely that major genes in category 4 (i.e., *GJB2* and *SLC26A4*) or their encoded proteins play a critical role during the embryonic stage of inner ear development, whereas major genes in category 3 (i.e., *MPZL2* and *KCNQ4*) or their encoded proteins act primarily in postnatal period. Collectively, these data support the development of a comprehensive genotype-phenotype map of SNHL, and shed light on insights of previously undefined genotype-phenotype correlations.

Discussion

Unlike previous cohort studies limited by the heterogeneous nature of phenotypes[33,34], our present findings provide a robust estimate of diagnostic rates through a stepwise genetic testing approach from single-gene analysis to WGS, integrated with functional assays and bioinformatic analysis, in a relatively large cohort of patients with a single phenotype of SNHL. Herein, we demonstrated an additional diagnostic yield of 19.2% (23/120) through the systematic application of WGS in previously undiagnosed SNHL patients who had undergone exome sequencing and targeted assays. Specifically, even among undiagnosed SNHL patients following an automated WGS pipeline (Step 3-1), we observed that the frequency of genome-wide single pathogenic allele in known recessive deafness genes was higher than in control cohorts. Based on these findings, we hypothesized that this elevated frequency of single pathogenic alleles (43%) is less likely to be incidental carrier variants and instead suggests the presence of a second, undetected hit in the opposite allele. Consequently, through SpliceAI-based deep intronic variants analysis (Step 3-2), we identified 3 meaningful deep intronic variants among 100 patients, including three *USH2A* deep intronic variants confirmed to be causative. These findings suggest the potential for further diagnosis of undiagnosed cases through additional molecular diagnostic approaches, such as multi-omics and methylation sequencing analyses, thereby accelerating the diagnostic process. Supporting this, Lunke et al. have shown the potential of the integration of multi-omic approaches into genomic testing, leading to additional diagnoses and changed critical care management[35].

We identified the genetic causes of SNHL in 55.6% of our cohort families, with WGS and SpliceAI-based deep intronic variants analysis increasing the overall

diagnostic yield by more than 5%. The improved diagnostic yield through WGS and in-depth analysis has been made possible by screening regions that are challenging to detect with conventional methods, primarily including deep intronic variants, small SVs, copy-neutral inversions, and complex genomic rearrangements. Among the 23 families with additional diagnoses identified through WGS and in-depth analysis, 16 cases (69.6%) could be diagnosed through reanalysis of targeted sequencing data and exome-based CNV algorithms. However, 7 (30.4%) of these cases required genome sequencing for a definitive diagnosis. Our study further provides clinical guidelines for selecting SNHL patients who are most likely to benefit from WGS. WGS proved more effective for genetic completion in SNHL patients with early-onset or syndromic features, regardless of audiological characteristics (e.g., severity and configuration), with trio-based WGS providing higher diagnostic yield.

Genetic diagnosis of SNHL is helpful not only for understanding clinical manifestations but also for planning treatment options. Genetic information could serve as a guide to clinical phenotypes and their natural course, highlighting the importance of WGS in identifying additional genetic causes in undiagnosed patients. In detail, early identification of genetic causes may be necessary for detecting preclinical symptoms (e.g., ns-SNHL mimics)[30], and for providing reproductive counseling, including guidance on next-baby planning and options for preimplantation genetic testing, even in cases of nonsyndromic hearing loss[36]. Although targeted agents based on genotype are not yet commonly established in many human genetic disorders, there have been significant advances in personalized targeted therapy in recent years, including in the field of genetic hearing loss[37-41]. For example, three presently identified genomic variants in the deep intronic region of *USH2A* are targetable by splice-switching antisense

oligonucleotide therapy (ASOs), offering an opportunity to slow down or even halt disease progression in these patients. According to a framework for individualized splice-switching ASO therapy[42], the *USH2A* deep intronic variants that induce pseudoexon inclusion without disrupting cryptic splicing sites were highly amenable to ASO splice modulation. Furthermore, in theory, a subset of SVs—such as the *EYA1* paracentric inversion (SNUH 734) and *EYA1* complex genomic rearrangements (SNUH 536) linked to haploinsufficiency detected in WGS—can be corrected using CRISPR-based editing approaches, including Cas9 nuclease with paired gRNAs, CRISPR activation, and prime editing strategy[43-45]. The present study provides good examples of the potential of inner ear precision medicine for SNHL treatment, with broadened therapeutic targets identified through WGS and in-depth analysis.

Our comprehensive genomic investigation further refined the genotype-phenotype landscape of SNHL, revealing gene signatures based on phenotypes. The distribution of phenotype-based signature genes identified in this study largely aligns with findings from exome-based SNHL cohort studies. While this genetic information could support genetic diagnosis and provide a rationale for in-depth analysis in the clinical setting, further studies with larger cohorts are essential to establish more specific phenotype-genotype correlations that account for variant effects (e.g., allelic hierarchy). Furthermore, we found that genotype-phenotype correlations were also present at the level of the molecular pathways of the genes in the inner ear. The inner ear molecular functions within distinct subcategories of genetic hearing loss were found to correlate with the single-cell expression patterns of identified deafness genes. This new framework of genetic hearing loss suggests that genes within each functional class exhibit not only distinct inner ear molecular

functions but also relatively homogeneous spatial expression patterns in the cochlea. Unlike traditional genotype-phenotype correlations, which are often limited by specific ethnicities, genotypes, or phenotypes, the classification based on inner ear molecular pathways expands beyond the current understanding of SNHL and provides insights for predicting phenotypes associated with newly identified deafness genes.

Collectively, our results provide evidence for the clinical utility of the integrated diagnostic approaches, including WGS, in real-world SNHL practice to fully recognize the genomic architectures and associated phenotypic attributes, paving the way toward precision medicine to come.

Methods

Study cohort

In this study, we utilized a prospective research design and focused on participants attending the Hereditary Hearing Loss Clinic within the Otorhinolaryngology division of the Center for Rare Diseases, Seoul National University Hospital, Korea, between March 2021 and February 2023 (**Fig. 1**). Patients were not included if they were referred from other centers with confirmed genetic diagnoses or diagnosed as conductive hearing loss. In total, our SNHL cohort comprised 394 unrelated families and 750 individuals including probands, who exhibited hearing loss with sensorineural components, and their family members. The demographic data and clinical phenotypes were retrieved from the electronic medical records. The onset of hearing loss was classified into three distinct categories[46]: early identification (i.e., congenital or prelingual deafness identified through failed newborn hearing screening test), delay identification (i.e., pediatric-onset deafness occurring by age 18 that does not meet the criteria for early identification, regardless of newborn hearing screening confirmation), and adult identification (i.e., documented adult-onset hearing loss). The syndromic features of the patients in the cohort were evaluated during their first outpatient clinic visit based on their medical histories and/or features in their clinical manifestations. The presence of associated medical conditions (e.g., syndromic hearing loss) was determined using the Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes. All procedures were approved by the Institutional Review Board of Seoul National University Hospital (no. IRB-H-0905-041-281 and IRB-H-2202-045-1298).

Audiological evaluation

Depending on the participant's age, the hearing thresholds for six different octaves (0.25, 0.5, 1, 2, 4, and 8 kHz) were evaluated using pure-tone audiometry (PTA)[47]. For patients under 3 years of age or having neurodevelopmental delay, auditory brainstem response threshold (ABRT) and auditory steady-state response (ASSR) were used to gauge the thresholds at four-octave frequencies (0.5, 1, 2, and 4 kHz). The conductive components were evaluated using comprehensive tests, including tympanic membrane examination, tympanometry (probe tones of 226 and 1000 Hz), and/or bone conduction ABRT, particularly in younger subjects. Auditory profiles were retrieved—such as asymmetry, severity, configuration, and progression. The mean hearing threshold was determined using an average of the thresholds at 0.5, 1, 2, and 4 kHz, and the degree of hearing loss was categorized as mild-to-moderate (21-40 dB or ≤ 20 dB with high-frequency hearing loss), moderate-to-severe (41-70 dB, and severe-to-profound (≥ 71 dB) [47]. Audiogram configurations were categorized into one of five subtypes: down-sloping (i.e., consistent downward trend observed across 250, 500, 1000, 2000, and 4000 Hz frequencies, with an average threshold at 250 and 500 Hz ≤ 40 dB), ski-sloping (i.e., thresholds at 250 Hz are ≤ 25 dB, with a decrease of ≥ 40 dB between 250–1 kHz or 500–2 kHz, or a decrease of ≥ 70 dB across 250–4 or 8 kHz), cookie-bite (i.e., U-shaped), up-sloping (i.e., rising), and flat (i.e., audiograms that does not fit down-sloping, ski-slope, cookie-bite, or up-sloping configurations)[48]. Asymmetric hearing loss was defined as severe-to-profound hearing loss in the poorer ear, with an average hearing threshold >30 dB HL and <55 dB HL in the better ear. The presence of interaural asymmetry (a difference in average between the poorer ear and the better ear of 15 to less than 30 dB, and a difference in average between the poorer ear and the better ear of 30 or

more than 30 dB) was also assessed[49]. To analyze hearing loss progression, serial audiograms were used to retrieve the hearing threshold at all frequencies. Hearing loss progression was assessed in cases with two or more audiograms documented during the follow-up period, with at least a one-year interval between documentation. Cases with only one audiogram or follow-up duration of less than 1-year were classified as not available (i.e., N/A). Profound SNHL with thresholds ≥90 dB at 500 Hz was classified as not determined (i.e., N/D). Hearing loss progression in this study was categorized as substantial (≥10 dB deterioration at three or more frequencies), mild (≥5 dB deterioration at three or more frequencies or ≥10 dB deterioration at one or two frequencies), and none (if neither substantial nor mild criteria were met).

Real-time polymerase chain reaction and *GJB2* sequencing

Genomic DNA was extracted from peripheral blood samples utilizing the Chemagic 360 instrument (Perkin Elmer, Baesweiler, Germany). Real-time polymerase chain reaction (PCR) was performed using the U-TOP™ HL Genotyping Kit Ver1 and Ver2, along with a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA)[50,51]. This process was used to examine 22 pathogenic variants across 10 deafness genes. The data collected from this procedure were analyzed using Bio-Rad CFX manager v1.6 software. Variants were identified through the fluorescence signals from the detection probes, which corresponded to the melting temperature (Tm), as specified by the standard protocol in the manufacturer's manual. We additionally conducted sequencing of the *GJB2* single gene, following a previously described method[52].

493 **Targeted panel sequencing and whole-exome sequencing**

494 We utilized TPS or WES to sequence the exonic regions of SNHL-related
495 genes. The target regions were captured using a SureSelect DNA targeted
496 sequencing panel for TPS, and a SureSelectXT Human All Exon V5 for WES (Agilent
497 Technologies, Santa Clara, CA, USA). A library was prepared following the
498 manufacturer's instructions, and was paired-end sequenced using a NovaSeq 6000
499 sequencing system (Illumina, San Diego, CA, USA).

500 Sequence reads were aligned to the human reference genome (GRCh38)
501 and processed according to the Genome Analysis Toolkit (GATK) best-practice
502 pipeline for calling single nucleotide variants (SNVs) and short insertions/deletions
503 (indels)[53]. The ANNOVAR program was used for variant annotation, such as the
504 RefSeq gene set and Genome Aggregation Database (gnomAD)[23,54]. Rare non-silent
505 variants were selected as candidates, including nonsynonymous SNVs, coding
506 indels, and splicing variants. We also used the Korean Reference Genome Database
507 (KRGDB) and KOVA databases for further filtration of ethnic-specific variants[24,55].
508 Additionally, the ClinVar and HGMD databases were screened to check whether
509 candidate variants had been previously identified in other patients[56,57].

510 We classified candidate variants according to the ACMG-AMP guidelines
511 using the InterVar program[26,58], and manually curated the classifications following the
512 modified guidelines for SNHL[59].

513

514 **Multiplex ligation-dependent probe amplification and mitochondria panel**
515 **sequencing**

516 For individuals displaying non-syndromic, symmetric, mild-to-moderate
517 SNHL, we evaluated copy number variations (CNVs) using the SALSA MLPA

Probemix P461-B1 *STRC-CATSPER2-OTOA* (MRC-Holland, Amsterdam, Netherlands)[21]. Additionally, for patients showing evidence of EVA on temporal bone CT and/or internal acoustic canal MRI and clinical features of BOR/BO syndrome, we performed *SLC26A4* and *EYA1* MLPA tests, respectively, using the SALSA MLPA Probemix P280-B4 *SLC26A4* and the SALSA MLPA Probemix P153-B2 *EYA1* (MRC-Holland). We analyzed the amplification products using an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and interpreted the results using Gene Marker 1.91 software (SoftGenetics, State College, PA, USA).

For mitochondria panel sequencing, DNA was extracted from peripheral blood samples using the Chemagic 360 instrument (Perkin Elmer, Baesweiler, Germany). The complete human mitochondrial genome was amplified in two overlapping fragments: fragment I (spanning 9,289 bp), and fragment II (spanning 7,626 bp). Fragment 1 was amplified using the primer pair hmtF1 569 (5′-AACCAAACCCCAAAGACACC-3′) and hmtR1 9819 (5′-GCCAATAATGACGTGAAGTCC-3′), and fragment II was amplified using the primer pair htmF2 9611 (5′-TCCCACTCCTAAACACATCC-3′) and hmtR2 626 (5′-TTTATGGGGTGATGTGAGCC-3′)[60]. PCR reactions were conducted using the following cycling parameters: initial denaturation at 94 °C for 2 min; 10 cycles of 94 °C for 15 s, 65 °C for 30 s, and 68 °C for 5 min; 25 cycles of 94 °C for 15 s, 65 °C for 30 s, and 68 °C for 5 min; and a final extension at 68 °C for 7 min. Subsequently, a library was generated using the Nextera DNA Flex Library Prep Kit (Illumina) following the manufacturer's instructions. Paired-end sequencing was performed with the generation of 150-bp reads on the MiSeq platform (Illumina). Bioinformatic processes, including alignment and annotation, were performed using NextGene Version 2.4.0.1 (Softgenetics).

Selection of the target population for whole-genome sequencing

All patients with s-SNHL (*n* = 21) who remained undiagnosed after exome sequencing and other techniques underwent WGS. Conversely, in patients with ns-SNHL (*n* = 177) who remained undiagnosed, we determined the sample representativeness for WGS. First, we estimated the sample size with a 7% margin of error and a 95% CI. Second, we employed a probability sampling method, specifically stratified sampling, considering a significant heterogeneity of SNHL patients with respect to audiological characteristics. Relevant covariates, including SNHL onset, severity, and asymmetry phenotypes, were used as criteria for stratification. Thus, for the 177 undiagnosed ns-SNHL patients, a representative validation was conducted, and WGS was ultimately performed on 99 families, exceeding the required sample size of 94, which corresponds to a 7% margin of error and a 95% CI. A representative sample was then obtained by randomly sampling within each stratum. Chi-square tests were conducted to assess differences across these strata. Overall, our methodology combines a well-thought-out sample size estimation with a stratified sampling approach and proper statistical validation, making it a robust approach for selecting a representative sample for WGS in undiagnosed ns-SNHL patients.

Library construction and automated analytic pipeline for whole-genome sequencing

To obtain genomic DNA, peripheral blood samples were collected from probands with or without their parents. The entire process of genome sequencing, analysis, and interpretation was performed using the RareVision™ system (Inocras,

San Diego, CA, USA). Genomic DNA was extracted from blood samples using the Allprep DNA/RNA kits (Qiagen, Venlo, Netherlands). DNA libraries were prepared using TruSeq DNA PCR-Free Library Prep Kits (Illumina) and sequenced on the Illumina NovaSeq6000 platform with an average depth of coverage of 30×. The obtained genome sequences were aligned to the human reference genome (GRCh38) using the BWA-MEM algorithm. PCR duplicates were removed using SAMBLASTER[61]. The initial mutation calling for base substitutions and short indels was performed using HaplotypeCaller and Strelka2, respectively[62]. SVs were identified using Delly. Variants were filtered, and their Mendelian inheritance patterns were assessed. De novo mutations were detected, and their potential impacts were predicted. The pathogenicity prediction was further enhanced by using in-house-developed software that automatically integrates updated databases. The final evaluation of variant pathogenicity was determined by medical geneticists, considering the patient's phenotype and familial history.

In vitro splicing analysis using minigene assay

Fragments carrying the *USH2A* intron 37 reduced to 916 bp containing c.7120+1475A or c.7120+1475A>G, and intron 64 reduced to 845 bp containing c.14135-3169A or c.14135-3169A>G, were amplified and cloned into the pSPL3 vector, between the exon splice donor (SD) and splice acceptor (SA), using the EcoRI and NdeI restriction sites. Human epithelial kidney 293T (HEK293T) cells were seeded in a six-well culture plate and incubated at 37□°C in a 5% CO_2 atmosphere in Dulbecco's modified Eagle's medium (LM001-05; Welgene, Gyeongsan, Korea) containing 10% fetal bovine serum (12483-020; Gibco, Carlsbad, CA, USA), 100 units/mL penicillin/streptomycin (LS015-01; Welgene), and 2 mM L-

593 glutamine (LS002-01; Welgene). On the next day, the cells were transfected with 2
594 µg pSPL3 plasmid using Lipofectamine 3000 reagent (L3000001; Invitrogen,
595 Carlsbad, CA, USA), according to the manufacturer's guidelines. After 24 hours, the
596 cells were harvested, and the total RNA was extracted using TRIzol Reagent
597 (15596026; Invitrogen) and chloroform. From 1 µg RNA, cDNA was prepared by
598 reverse transcription using the Accupower RT-preMix (K-2041; Bioneer, Daejeon,
599 Korea). Splicing analysis was performed by PCR amplification with Taq DNA
600 Polymerase (E-2011-1; Bioneer) using the following vector-specific primers: SD6
601 (5□-TCTGAGTCACCTGGACAACC-3□) and SA2 (5□-
602 ATCTCAGTGGTATTTGTGAGC-3□).
603
604 **Fibroblast cell culture**
605 A skin biopsy was obtained from a donor under local anesthesia and
606 preserved in Phosphate Buffered Saline (PBS). The biopsy was divided into 9-12
607 distinct segments and seeded into a 12-well plate containing DMEM supplemented
608 with 20% FBS. Once the segments reached confluence, fibroblasts were harvested
609 for further expansion.
610
611 **Oxygen consumption rate**
612 Cellular OCR was measured in real-time using the Seahorse XF96
613 Extracellular Flux Analyzer (Seahorse Bioscience, North Billerica, MA, USA) per the
614 manufacturer's protocol. Cells (8.0 × 10^3 fibroblasts) were seeded in 100 µL of
615 growth medium in Seahorse 96-well microplates and incubated at 37 ºC with 5%
616 CO2 for 24 hours. Prior to the assay, cells were washed with assay running media
617 (unbuffered DMEM supplemented with 25 mM glucose, 1 mM glutamine, and 1 mM

sodium pyruvate) and equilibrated in a non-CO2 incubator overnight. Calibration of the assay plate was performed overnight in a non-CO2 incubator. Once calibrated, the cell plate replaced the assay plate, and OCR was measured simultaneously. The assay protocol involved sequential injection of four compounds to modulate mitochondrial function and determine parameters such as basal respiration, maximal respiration, and ATP production: oligomycin (1 µM), an ATP synthase inhibitor for maximal glycolytic metabolism; carbonyl cyanide p-(trifluoromethoxy) phenylhydrazone (FCCP) (1 µM), an ETC and OXPHOS uncoupler for peak oxygen consumption and oxidative metabolism; rotenone and antimycin A (both at 1 µM), inhibitors of ETC complexes I and III respectively, for non-mitochondrial respiration assessment. The Seahorse analyzer recorded OCR values throughout the assay to monitor cellular metabolic activity in real-time.

Statistical analyses

We used the Pearson chi-square test to identify variables that could potentially differentiate between the genetically diagnosed and undiagnosed groups. Following the computation of the OR and 95% CI for each value, we conducted a logistic regression analysis, considering only the variables with *P* values of <0.05. This approach facilitated the derivation of the adjusted OR. We also classified the 63 SNHL-associated genes that were identified during this study based on the molecular mechanisms of inner-ear function. To verify the variance between groups according to clinical phenotypes, we used the Pearson chi-square test. Finally, we used Fisher's exact test based on the false discovery rate (FDR) to identify statistically significant groups characterized by an adjusted *P* value of <0.05. In this study, we define variables exhibiting *P* values of <0.05 as statistically significant.

Statistical analyses and visualizations were conducted using R Version 4.2.2, and the corresponding code can be accessed at https://github.com/SNUH-hEARgeneLab/WGS_analysis.

Data availability:

The data generated in this study are available in the Source Data file, which is provided with this paper.

References

1 Lieu, J. E. C., Kenna, M., Anne, S. & Davidson, L. Hearing Loss in Children: A Review. *JAMA* **324**, 2195-2205 (2020). https://doi.org/10.1001/jama.2020.17647

2 Bowl, M. R. *et al.* A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nature communications* **8**, 886 (2017).

3 Pinero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* **48**, D845-D855 (2020). https://doi.org/10.1093/nar/gkz1021

4 Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* **51**, 1560-1565 (2019). https://doi.org/10.1038/s41588-019-0528-2

5 Adams, D. R. & Eng, C. M. Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. *N Engl J Med* **379**, 1353-1362 (2018). https://doi.org/10.1056/NEJMra1711801

6 Pajusalu, S. *et al.* Large gene panel sequencing in clinical diagnostics-results from 501 consecutive cases. *Clin Genet* **93**, 78-83 (2018). https://doi.org/10.1111/cge.13031

7 Aspromonte, M. C. *et al.* Characterization of intellectual disability and autism comorbidity through gene panel sequencing. *Hum Mutat* **40**, 1346-1363 (2019). https://doi.org/10.1002/humu.23822

8 Liao, E. N., Taketa, E., Mohamad, N. I. & Chan, D. K. Outcomes of Gene Panel Testing for Sensorineural Hearing Loss in a Diverse Patient Cohort. *JAMA Netw Open* **5**, e2233441 (2022). https://doi.org/10.1001/jamanetworkopen.2022.33441

9 Downie, L. *et al.* Exome sequencing in infants with congenital hearing impairment: a population-based cohort study. *Eur J Hum Genet* **28**, 587-596 (2020). https://doi.org/10.1038/s41431-019-0553-8

10 Gu, X. *et al.* Genetic testing for sporadic hearing loss using targeted massively parallel sequencing identifies 10 novel mutations. *Clin Genet* **87**, 588-593 (2015). https://doi.org/10.1111/cge.12431

11 Jung, J. *et al.* Genetic Predisposition to Sporadic Congenital Hearing Loss in

a Pediatric Population. *Sci Rep* **7**, 45973 (2017). https://doi.org/10.1038/srep45973

12 Pennisi, E. Upstart DNA sequencers could be a “game changer.”. *Science* **376**, 1257-1258 (2022).

13 Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine* **18**, 696-704 (2016).

14 Wojcik, M. H. *et al.* Genome Sequencing for Diagnosing Rare Diseases. *N Engl J Med* **390**, 1985-1997 (2024). https://doi.org/10.1056/NEJMoa2314761

15 Lee, S. *et al.* Whole genomic approach in mutation discovery of infantile spasms patients. *Front Neurol* **13**, 944905 (2022). https://doi.org/10.3389/fneur.2022.944905

16 Gallego-Martinez, A. *et al.* Using coding and non-coding rare variants to target candidate genes in patients with severe tinnitus. *NPJ Genom Med* **7**, 70 (2022). https://doi.org/10.1038/s41525-022-00341-w

17 Huq, A. J. *et al.* Clinical impact of whole-genome sequencing in patients with early-onset dementia. *J Neurol Neurosurg Psychiatry* (2022). https://doi.org/10.1136/jnnp-2021-328146

18 Macken, W. L. *et al.* Specialist multidisciplinary input maximises rare disease diagnoses from whole genome sequencing. *Nat Commun* **13**, 6324 (2022). https://doi.org/10.1038/s41467-022-32908-7

19 van Eyk, C. L. *et al.* Yield of clinically reportable genetic variants in unselected cerebral palsy by whole genome sequencing. *NPJ Genom Med* **6**, 74 (2021). https://doi.org/10.1038/s41525-021-00238-0

20 Pagnamenta, A. T. *et al.* Structural and non-coding variants increase the diagnostic yield of clinical whole genome sequencing for rare diseases. *Genome Med* **15**, 94 (2023). https://doi.org/10.1186/s13073-023-01240-0

21 Kim, B. J. *et al.* Significant Mendelian genetic contribution to pediatric mild-to-moderate hearing loss and its comprehensive diagnostic approach. *Genetics in Medicine* **22**, 1119-1128 (2020).

22 Nam, D. W. *et al.* Molecular Genetic Etiology and Revisiting the Middle Ear Surgery Outcomes of Branchio-Oto-Renal Syndrome: Experience in a Tertiary Referral Center. *Otology & Neurotology* **44**, e319-e327 (2023).

23 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from

variation in 141,456 humans. *Nature* **581**, 434-443 (2020). https://doi.org/10.1038/s41586-020-2308-7

24 Lee, J. *et al.* A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. *Exp Mol Med* **54**, 1862-1871 (2022). https://doi.org/10.1038/s12276-022-00871-4

25 Choi, W. H. *et al.* Functional pathogenicity of ESRRB variant of uncertain significance contributes to hearing loss (DFNB35). *Sci Rep* **14**, 21215 (2024). https://doi.org/10.1038/s41598-024-70795-8

26 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015). https://doi.org/10.1038/gim.2015.30

27 Buchert, R. *et al.* SPATA5 mutations cause a distinct autosomal recessive phenotype of intellectual disability, hypotonia and hearing loss. *Orphanet Journal of Rare Diseases* **11**, 1-7 (2016).

28 Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548 e524 (2019). https://doi.org/10.1016/j.cell.2018.12.015

29 Yun, Y. *et al.* Expanding Genotype-Phenotype Correlation of CLCNKA and CLCNKB Variants Linked to Hearing Loss. *Int J Mol Sci* **24** (2023). https://doi.org/10.3390/ijms242317077

30 Nam, D. W. *et al.* Allelic hierarchy for USH2A influences auditory and visual phenotypes in South Korean patients. *Sci Rep* **13**, 20239 (2023). https://doi.org/10.1038/s41598-023-47166-w

31 Delmaghani, S. & El-Amraoui, A. Inner ear gene therapies take off: current promises and future challenges. *Journal of clinical medicine* **9**, 2309 (2020).

32 van der Valk, W. H. *et al.* A single-cell level comparison of human inner ear organoids with the human cochlea and vestibular organs. *Cell Rep* **42**, 112623 (2023). https://doi.org/10.1016/j.celrep.2023.112623

33 Investigators, G. P. P. *et al.* 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* **385**, 1868-1880 (2021). https://doi.org/10.1056/NEJMoa2035790

34 Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a

national health system. *Nature* **583**, 96-102 (2020). https://doi.org/10.1038/s41586-020-2434-2

35 Lunke, S. *et al.* Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat Med* **29**, 1681-1691 (2023). https://doi.org/10.1038/s41591-023-02401-9

36 Yang, Y. *et al.* Reevaluating the splice-altering variant in TECTA as a cause of nonsyndromic hearing loss DFNA8/12 by functional analysis of RNA. *Hum Mol Genet* (2024). https://doi.org/10.1093/hmg/ddae071

37 Jiang, L., Wang, D., He, Y. & Shu, Y. Advances in gene therapy hold promise for treating hereditary hearing loss. *Molecular Therapy* (2023).

38 Petit, C., Bonnet, C. & Safieddine, S. Deafness: from genetic architecture to gene therapy. *Nature Reviews Genetics*, 1-22 (2023).

39 Yun, Y. & Lee, S. Y. Updates on Genetic Hearing Loss: From Diagnosis to Targeted Therapies. *J Audiol Otol* **28**, 88-92 (2024). https://doi.org/10.7874/jao.2024.00157

40 Lv, J. *et al.* AAV1-hOTOF gene therapy for autosomal recessive deafness 9: a single-arm trial. *Lancet* **403**, 2317-2325 (2024). https://doi.org/10.1016/S0140-6736(23)02874-X

41 Wang, H. *et al.* Bilateral gene therapy in children with autosomal recessive deafness 9: single-arm trial results. *Nat Med* **30**, 1898-1904 (2024). https://doi.org/10.1038/s41591-024-03023-5

42 Kim, J. *et al.* A framework for individualized splice-switching oligonucleotide therapy. *Nature* **619**, 828-836 (2023). https://doi.org/10.1038/s41586-023-06277-0

43 Kweon, J. *et al.* Targeted genomic translocations and inversions generated using a paired prime editing strategy. *Molecular Therapy* **31**, 249-259 (2023).

44 Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nature Reviews Genetics* **24**, 161-177 (2023).

45 Yi, H. *et al.* CRISPR-based editing strategies to rectify EYA1 complex genomic rearrangement linked to haploinsufficiency. *Mol Ther Nucleic Acids* **35**, 102199 (2024). https://doi.org/10.1016/j.omtn.2024.102199

46 Liao, E. N., Taketa, E., Mohamad, N. I. & Chan, D. K. Outcomes of Gene Panel Testing for Sensorineural Hearing Loss in a Diverse Patient Cohort.

JAMA Network Open **5**, e2233441-e2233441 (2022).

47 Lee, S.-Y. *et al.* Natural course of residual hearing with reference to GJB2 and SLC26A4 genotypes: Clinical implications for hearing rehabilitation. *Ear and Hearing* **42**, 644-653 (2021).

48 Lee, S.-Y. *et al.* No auditory experience, no tinnitus: lessons from subjects with congenital-and acquired single-sided deafness. *Hearing research* **354**, 9-15 (2017).

49 Lin, P.-H. *et al.* Etiologic and audiologic characteristics of patients with pediatric-onset unilateral and asymmetric sensorineural hearing loss. *JAMA otolaryngology–head & neck surgery* **143**, 912-919 (2017).

50 Han, K.-H. *et al.* Establishment of a flexible real-time polymerase chain reaction-based platform for detecting prevalent deafness mutations associated with variable degree of sensorineural hearing loss in Koreans. *PloS one* **11**, e0161756 (2016).

51 Lee, S.-Y. *et al.* Flexible real-time polymerase chain reaction-based platforms for detecting deafness mutations in Koreans: A proposed guideline for the etiologic diagnosis of auditory neuropathy spectrum disorder. *Diagnostics* **10**, 672 (2020).

52 Kim, S. Y. *et al.* Prevalence of p. V37I variant of GJB2 in mild or moderate hearing loss in a pediatric population and the interpretation of its pathogenicity. *PloS one* **8**, e61592 (2013).

53 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-11 10 33 (2013). https://doi.org/10.1002/0471250953.bi1110s43

54 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010). https://doi.org/10.1093/nar/gkq603

55 Jung, K. S. *et al.* KRGDB: the large-scale variant database of 1722 Koreans based on whole genome sequencing. *Database (Oxford)* **2020** (2020). https://doi.org/10.1093/database/baaa030

56 Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research,

genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665-677 (2017). https://doi.org/10.1007/s00439-017-1779-6

57 Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062-D1067 (2018). https://doi.org/10.1093/nar/gkx1153

58 Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* **100**, 267-280 (2017). https://doi.org/10.1016/j.ajhg.2017.01.004

59 Oza, A. M. *et al.* Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat* **39**, 1593-1613 (2018). https://doi.org/10.1002/humu.23630

60 Tang, S. & Huang, T. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques* **48**, 287-296 (2010). https://doi.org/10.2144/000113389

61 Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505 (2014). https://doi.org/10.1093/bioinformatics/btu314

62 Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591-594 (2018). https://doi.org/10.1038/s41592-018-0051-x

Acknowledgements: This research was supported and funded by SNUH Kun-hee Lee Child Cancer & Rare Disease Project, Republic of Korea (grant number: FP-2022-00001-004 to Sang-Yeon Lee), SNU Medicine grant (basic and clinic cooperation research grant number: 800-20230428 to Sang-Yeon Lee), and National Research Foundation of Korea (NRF) and funded by the Ministry of Education (grant number: 2022R1C1C1003147 to Sang-Yeon Lee). This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI22C182600 to June-Young Koh).

Contributions: Concept and design: S.-Y.L., J.-H.C. Acquisition, analysis, or interpretation of data: S.-Y.L., S.L., S.P., J-.Y.K., S.H.J., Y.J., W.H.C., J.H.C. Drafting of the manuscript: S.-Y.L., S.L., S.P., J-.Y.K. Critical review of the manuscript for important intellectual content: S.-Y.L., S.L., S.P., J-.Y.K. Statistical analysis: S.-Y.L., S.P., J-.Y.K. Obtained funding: S.-Y.L., J-.Y.K. Administrative, technical, or material support: H.Y., S.L., M.-W.S., M.K.P., J.-J.S., B.Y.C., J.H.L. Supervision: J.-H.C. Y.S.J.

Competing interests:

Young Seok Ju is the founder of Inocras Inc., a genome analysis and interpretation company. Young Seok Ju, June-Young Koh, Seongyeol Park, and Sangmoon Lee hold stocks or stock options in Inocras Inc.

Figure legends

Figure 1. Study design and diagnostic pipeline. Flow diagram illustrating a prospective, step-by-step genetic approachof 394 unrelated SNHL families and 750 individuals including probands in our cohort study. The diagnostic pipeline included Step 1 (real-time PCR screening kit and direct *GJB2* sequencing), Step 2-1 (targeted panel sequencing or whole-exome sequencing), Step 2-2 (MLPA and/or mtDNA panel sequencing), Step 3-1 (whole-genome sequencing), and Step 3-2 (deep intronic variant analysis). SNHL, sensorineural hearing loss; BOR/BO syndrome, branchio-oto-renal/branchio-otic syndrome; MLPA, multiplex ligation-dependent probe amplification; mtDNA, mitochondrial DNA; CI, confidence interval.

Figure 2. Stepwise genetic diagnosis outcomes in SNHL patients. (a) Diagnostic yield of each genetic test for the whole SNHL cohort. Bar graph showing the cumulative diagnostic rate according to genetic diagnosis steps. (b-c) Diagnostic yield based on SNHL phenotypes and comparative analysis within the whole SNHL cohort (n = 394 families; b) and within the WGS cases (n = 120 families; c). Statistical significance for hearing loss onset and syndromic features was determined using one-way ANOVA with Bonferroni's multiple comparisons tests and the t-test, respectively. Significance levels are indicated as $*P < 0.05$. (d) Pearson's correlation coefficient values for causative variants between the allele frequencies (AFs) in our cohort and those from other populations. The AFs in our cohort showed the highest correlation with the Korean database, KOVA, followed by the East Asian population within gnomAD. (e) Dot plots shows the AF correlation between our cohort and KOVA. There are three variants (pink dots) showing higher AFs in KOVA, suggesting their phenotypic variability such as low penetrance or late onset

associated with these variants. A fitted line of linear regression model (blue line) and 95% confidence intervals (grey area) are displayed. (f) Distribution of variant subtypes identified at each diagnostic step. WGS, whole-genome sequencing; MLPA, multiplex ligation-dependent probe amplification; mtDNA, mitochondrial DNA; SNHL, sensorineural hearing loss; SNV, single nucleotide variant; indel, insertion/deletion.

Figure 3. Genomic landscape of the SNHL cohort.

(a) Bar plot showing the frequencies and inheritance patterns of 63 SNHL-associated genes from 219 genetically diagnosed families. Pie chart showing the percentages of inheritance patterns. (b) Bar plot showing the mutational landscape of the total 352 likely pathogenic or pathogenic variants among the 63 SNHL genes. Pie chart showing the percentages of variant types. (c) Proportion of novel variants among identified causal variants. (d) Structural variants (SVs) were more common among novel variants compared to previously reported variants. (e) Novel variants were frequently identified through WGS (Step 3-1) and SpliceAI-based deep intronic variants analysis (Step 3-2). (f, g) Schematic illustrations showing the location of each identified pathogenic variant within *SPATA5* and *CLCNKA-CLCNKB* in the probands (top, respectively). The genomic regions corresponding to each variant are visualized using the integrative genomics viewer (bottom) for each figure. SNHL, sensorineural hearing loss; SNV, single nucleotide variant; indel, insertion/deletion; SV, structural variation.

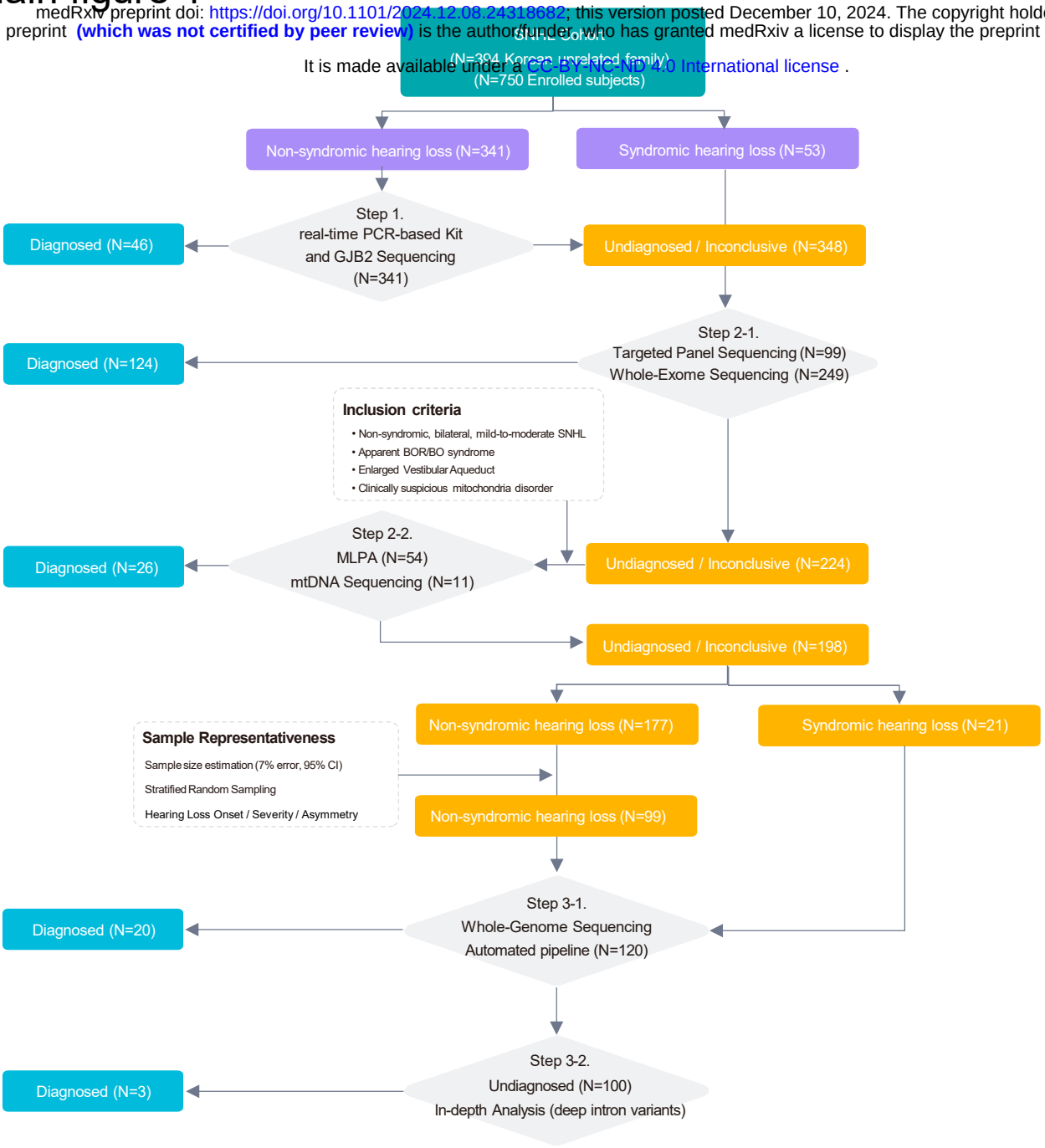
Figure 4. Analysis of carrier and deep intronic variants in undiagnosed SNHL patients.

(a) Schematic diagram illustrating carrier status identification and screening for

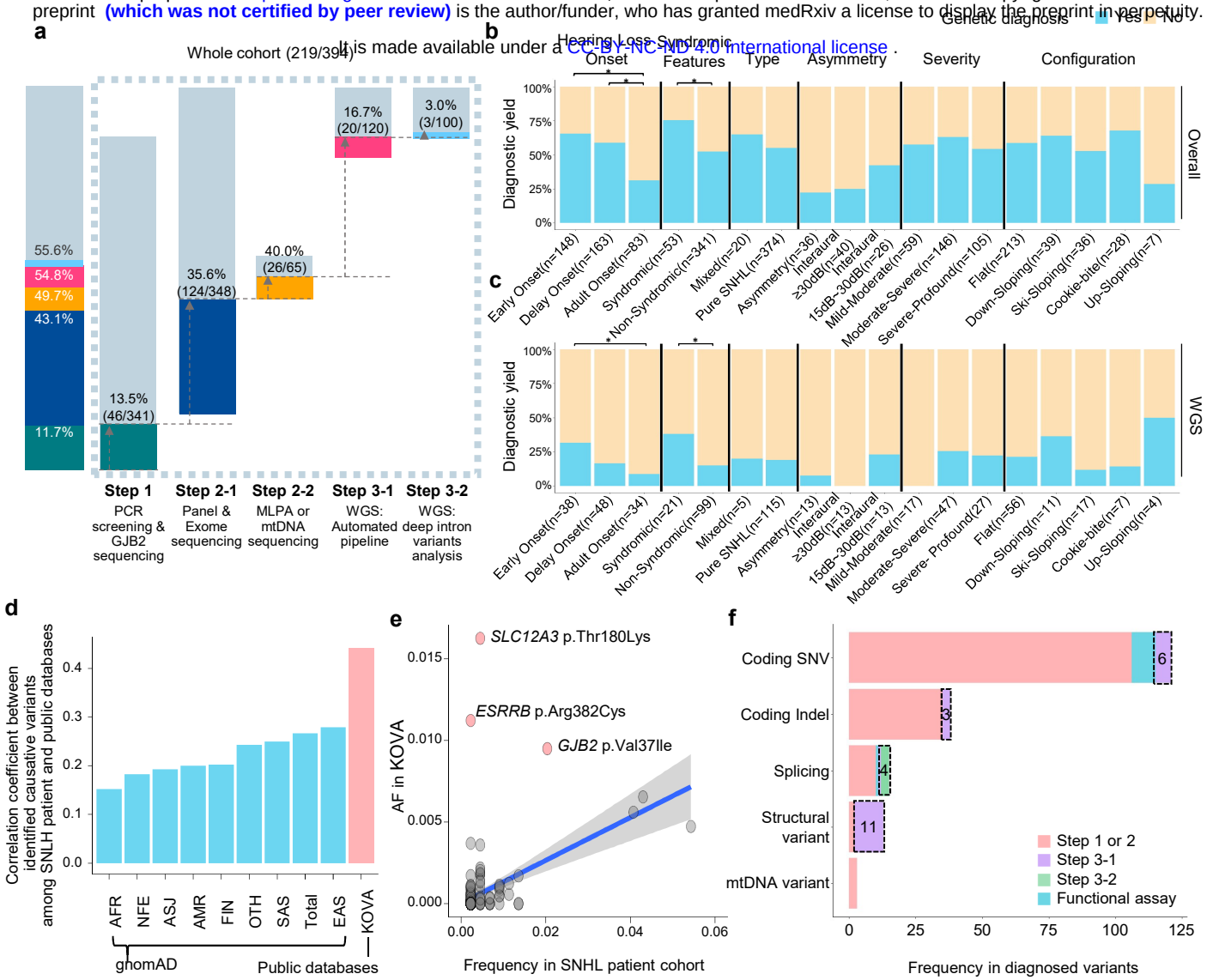
pathogenic deep intronic variants. (b) Bar plot showing the proportion of genetic carriers for SNHL-related genes across each cohort. (c) Cumulative distribution plot showing the MAF of candidate pathogenic variants across cohorts. (d) Schematic illustration of the location of each identified pathogenic variant within *USH2A* in each patient. (e) Schematic diagram of the pSPL3 vector with *USH2A* c.7120+1475A>G (left), c.14134-3169A>G (middle), and c.4628-26037A>G (right). (f) Electrophoresis gel image showing the bands corresponding to the pSPL3 empty vector (263 bp), each variant, and wild-type. (g) Schematic representation of the splice products with the wild-type splicing profile, and the splice variant profiles for each mutant type. SNHL, sensory neural hearing loss; WGS, whole-genome sequencing; HC, healthy control; AR, autosomal recessive; MOI, mode of inheritance; P, pathogenic; LP, likely pathogenic; MAF, minor allele frequency; SV, structural variation; TE, transposable element; SD, splicing donor; SA, splicing acceptor; PE, pseudoexon.

Figure 5. Functional classification and clinical relevance of SNHL-related genes. (a) Schematic illustration of seven functional categories according to the molecular mechanisms related to inner-ear function. (b) Heatmap showing the correlation of gene expression patterns across different gene categories. (c) Comparison of specific phenotypic presentations by seven inner ear functional categories. The x-axis represents classifications based on inner ear molecular functions, and y-axis indicates the proportion of clinical phenotypes among affected probands. (d) Bar graphs showing post-hoc analysis using the false discovery rate. Red represents the proportion of affected probands within the total number of probands assigned to specific inner ear molecular functional groups.* denotes statistical significance ($P < 0.05$).

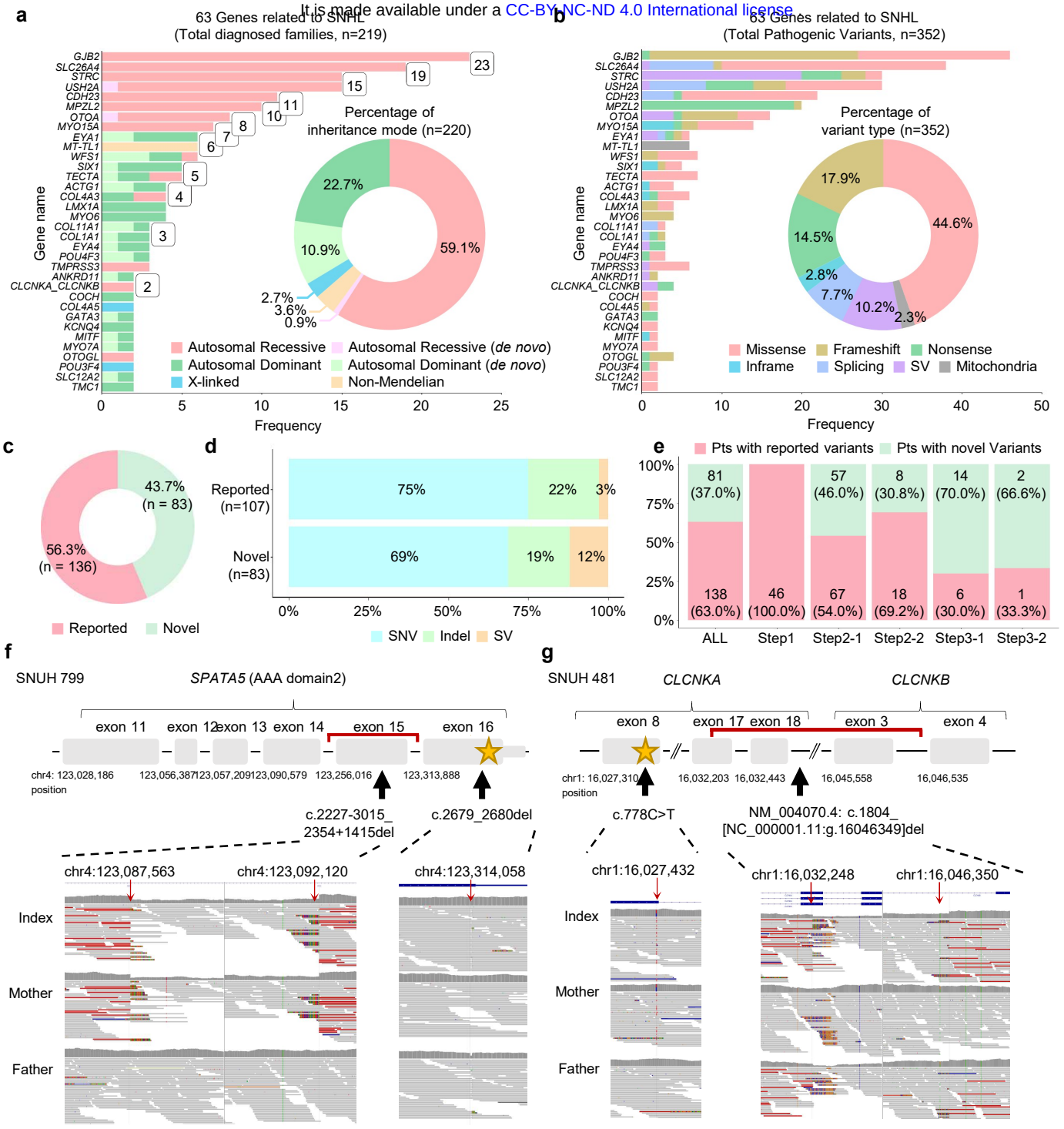
Main figure 1



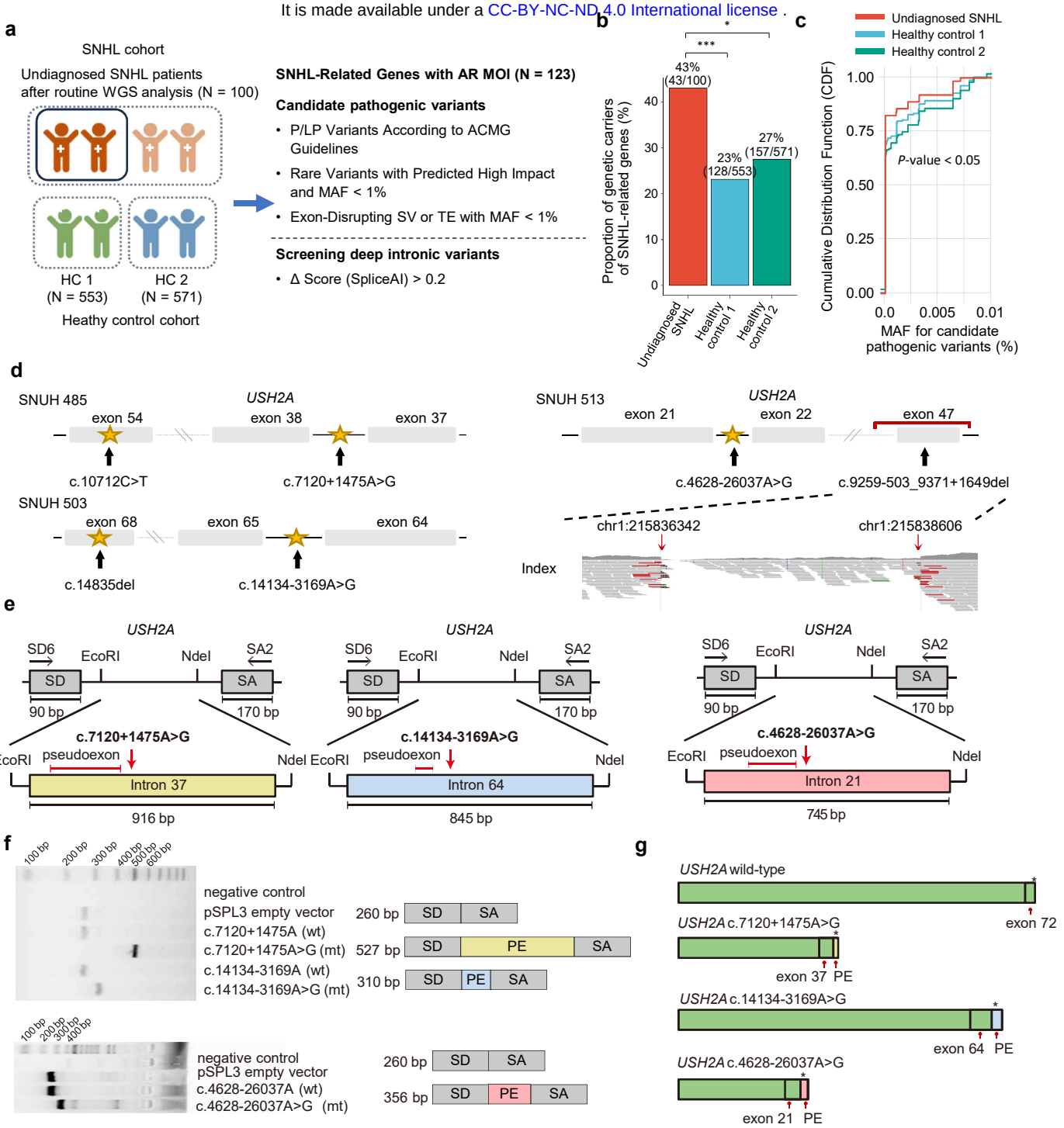
Main figure 2



Main figure 3



Main figure 4



Main figure 5