Utilizing Ethnic-Specific Differences in Minor Allele Frequency to Recategorize Reported Pathogenic Deafness Variants

A. Eliot Shearer,[1] Robert W. Eppsteiner,[1,18] Kevin T. Booth,[1,18] Sean S. Ephraim,[2,18] José Gurrola, II,[1] Allen Simpson,[1] E. Ann Black-Ziegelbein,[1] Swati Joshi,[3] Harini Ravi,[3] Angelica C. Giuffre,[3] Scott Happe,[3] Michael S. Hildebrand,[4] Hela Azaiez,[1] Yildirim A. Bayazit,[5] Mehmet Emin Erdal,[6] Jose A. Lopez-Escamez,[7] Irene Gazquez,[7] Marta L. Tamayo,[8] Nancy Y. Gelvez,[8] Greizy Lopez Leal,[8] Chaim Jalas,[9] Josef Ekstein,[10] Tao Yang,[11] Shin-ichi Usami,[12] Kimia Kahrizi,[13] Niloofar Bazazzadegan,[13] Hossein Najmabadi,[13] Todd E. Scheetz,[2,14,15] Terry A. Braun,[2,14,15] Thomas L. Casavant,[2,14,15] Emily M. LeProust,[3,19] and Richard J.H. Smith[1,16,17,*]

Ethnic-specific differences in minor allele frequency impact variant categorization for genetic screening of nonsyndromic hearing loss (NSHL) and other genetic disorders. We sought to evaluate all previously reported pathogenic NSHL variants in the context of a large number of controls from ethnically distinct populations sequenced with orthogonal massively parallel sequencing methods. We used HGMD, ClinVar, and dbSNP to generate a comprehensive list of reported pathogenic NSHL variants and re-evaluated these variants in the context of 8,595 individuals from 12 populations and 6 ethnically distinct major human evolutionary phylogenetic groups from three sources (Exome Variant Server, 1000 Genomes project, and a control set of individuals created for this study, the OtoDB). Of the 2,197 reported pathogenic deafness variants, 325 (14.8%) were present in at least one of the 8,595 controls, indicating a minor allele frequency (MAF) >0.00006. MAFs ranged as high as 0.72, a level incompatible with pathogenicity for a fully penetrant disease like NSHL. Based on these data, we established MAF thresholds of 0.005 for autosomal-recessive variants (excluding specific variants in *GJB2*) and 0.0005 for autosomal-dominant variants. Using these thresholds, we recategorized 93 (4.2%) of reported pathogenic variants as benign. Our data show that evaluation of reported pathogenic deafness variants using variant MAFs from multiple distinct ethnicities and sequenced by orthogonal methods provides a powerful filter for determining pathogenicity. The proposed MAF thresholds will facilitate clinical interpretation of variants identified in genetic testing for NSHL. All data are publicly available to facilitate interpretation of genetic variants causing deafness.

The advent of massively parallel sequencing has shifted the bottleneck in human genetics from data acquisition to variant interpretation. Accurate evaluation of genetic variants for pathogenicity is crucial to advance our understanding of disease processes and is a requirement for clinical diagnostics. Whole-exome sequencing and targeted gene panels based on targeted genomic enrichment and massively parallel sequencing are becoming commonplace and for some Mendelian diseases, including breast and ovarian cancer, degenerative eye disease, and hearing loss, they have become the ideal test and are now used routinely for clinical diagnostic testing.[1–3] These tests regularly produce hundreds or thousands of variants that require categorization and interpretation to assess their likelihood of causing disease. Correct interpretation is crucial when test results are used to direct clinical care.

Hearing loss (HL) is the most common sensory deficit in humans, affecting 1 in 500 children[4] and 360 million people worldwide (World Health Organization Deafness Estimate Online Report). The majority of HL is genetic and nonsyndromic (NSHL, not associated with other clinical phenotypes). Genetic diagnosis for NSHL is particularly challenging given limited phenotypic variability and extreme genetic heterogeneity; increased use of massively parallel sequencing resulting in thousands of variants identified per individual has highlighted these challenges.

[1]Molecular Otolaryngology & Renal Research Labs, Department of Otolaryngology—Head and Neck Surgery, University of Iowa Carver College of Medicine, Iowa City, IA 52242, USA; [2]Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242, USA; [3]Agilent Technologies, Cedar Creek, TX 78612, USA; [4]Epilepsy Research Centre, Department of Medicine, University of Melbourne, Heidelberg, VIC 3084, Australia; [5]Department of Otolaryngology, Faculty of Medicine, Medipol University, Istanbul 34083, Turkey; [6]Department of Medical Biology and Genetics, University of Mersin, Mersin 33160, Turkey; [7]Otology and Neurotology Group CTS495, Center for Genomic and Oncological Research (GENyO), Granada 18012, Spain; [8]Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá 11001000, Colombia; [9]Bonei Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, NY 11204, USA; [10]Dor Yeshorim, The Committee for Prevention of Jewish Genetic Diseases, Brooklyn, NY 11211, USA; [11]Department of Otorhinolaryngology—Head and Neck Surgery, Xinhua Hospital, and the Ear Institute, Shanghai Jiaotong University School of Medicine, Shanghai 20025, China; [12]Department of Otorhinolaryngology, School of Medicine, Shinshu University, Matsumoto, Nagano 390-8621, Japan; [13]Genetics Research Centre, University of Social Welfare and Rehabilitation Sciences, Tehran 1985713834, Iran; [14]Center for Bioinformatics and Computational Biology, University of Iowa, Iowa City, IA 52242, USA; [15]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242, USA; [16]Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA 52242, USA; [17]Department of Molecular Physiology & Biophysics, University of Iowa Carver College of Medicine, Iowa City, IA 52242, USA
[18]These authors contributed equally to this work
[19]Present address: Twist Biosciences, San Francisco, CA 94158, USA
*Correspondence: richard-smith@uiowa.edu
http://dx.doi.org/10.1016/j.ajhg.2014.09.001.

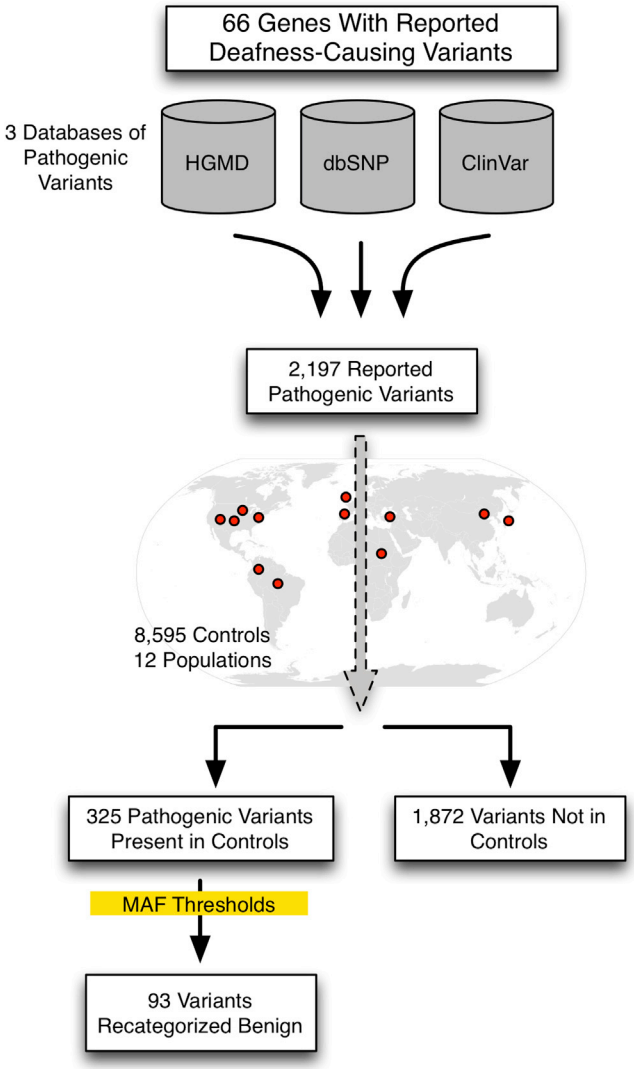

Figure 1. Study Overview

To date more than 70 genes and more than 2,000 causal variants have been implicated in NSHL. As with other genetic diseases, the vast majority of reported pathogenic NSHL variants were so designated based on five primary criteria: (1) cosegregation with the phenotype ideally with linkage analysis, (2) absence in 200 or more control chromosomes from the same ethnicity, (3) conservation of the affected nucleotide or amino acid through evolution, (4) predicted functional effect of the mutation, and (5) corroborative in vitro or in vivo functional data.

Absent, however, has been a sixth criterion—one that would require assessment of the variant in the context of a large number (thousands) of controls from varying ethnicities. This omission is noteworthy because the majority of human genetic variation is accounted for by ethnicity-specific differences.[5,6] Our group and others have hypothesized that evaluation of variants in this context can reduce or eliminate false positives. Massively parallel sequencing techniques have for the first time delivered the ability to sequence a heretofore-unobtainable number of control individuals from different ethnicities with data freely available to the scientific community.

In the first application of a large amount of control data to evaluate a human genetic disease, Norton et al. proposed a disease-specific iterative approach: use a large number of controls to determine the minor allele frequency (MAF) of known pathogenic variants and thereby determine a MAF cut-off that can be used for evaluation of future variants.[7] The authors used data from the Exome Variant Server (EVS), which includes two ethnically distinct populations, to resolve false positives. Another recent study disqualified ten reported X-linked disability genes using similar methods and data from the EVS.[8] Most recently, a gene with variants reported to cause NSHL, *MYO1A* (MIM 601478), has been disqualified through the use of EVS and lack of cosegregation of previously reported pathogenic variants in this gene.[9]

Incorrect classification of genetic variants as pathogenic is not limited to only one or a few genetic diseases—it is a systemic issue in human genetics. Cassa et al. evaluated all of the mutations present in the Human Gene Mutation Database (HGMD), the most comprehensive repository of pathogenic human mutations, in the context of control data from the 1000 Genomes Project.[10] The authors found that 3.5% of variants present in HGMD have a MAF > 0.05, which is implausible for a highly penetrant Mendelian disease.

Our goal in this study was to evaluate all reported pathogenic deafness variants in the context of a large number of controls from multiple ethnicities in order to recategorize any clearly benign variants based on MAF in controls and establish MAF thresholds to aid interpretation of variants discovered in the future (Figure 1).

We sought to evaluate pathogenic variants present in 66 genes with variants reported to cause NSHL and NSHL-mimic syndromes (genetic causes of syndromic hearing loss that mimic NSHL at presentation such as Usher syndrome, the most common cause of deaf-blindness; see Table S1 [available online] for a complete gene list). Reported pathogenic deafness variants were obtained from three sources: HGMD (Professional v.2013.4), NCBI ClinVar, and dbSNP v.138 (the latter publically available databases were accessed in February 2014). When necessary, variants were converted to HGVS nomenclature using Mutalyzer to generate genomic coordinates.[11] The final comprehensive list of reported pathogenic deafness variants included all variants categorized as fully penetrant pathogenic or probable-pathogenic in any of the databases (DM or DM? in the case of HGMD).

In total, we obtained 2,197 variants designated as pathogenic or probable-pathogenic by any of three databases. As shown in Figure S1, the majority of these variants were identified in HGMD alone (1,572; 71.6%), seven were unique to ClinVar, and none were unique to dbSNP. The three databases shared 454 (20.7%) of the reported pathogenic variants.