Published in final edited form as:

Laryngoscope. 2009 November ; 119(11): 2211–2215. doi:10.1002/lary.20664.

AudioGene Audioprofiling: A Machine-based Candidate Gene Prediction Tool for Autosomal Dominant Non-syndromic Hearing Loss

Michael S. Hildebrand, PhD^{a,#}, Adam P. DeLuca^{b,c,#}, Kyle R. Taylor^{b,c}, David P. Hoskinson^c, In Ae Hur^{b,c}, Dylan Tack^c, Sarah J. McMordie, MS^{a,d}, Patrick L.M. Huygen, PhD^e, Thomas L. Casavant, PhD^{b,c,f,g,#}, and Richard J.H. Smith, MD^{a,d,#}

^aDepartment of Otolaryngology - Head and Neck Surgery, University of Iowa, Iowa City, IA 52242, USA ^bDepartment of Biomedical Engineering, The University of Iowa, Iowa City, Iowa 52242, USA ^cCenter for Bioinformatics and Computational Biology, The University of Iowa, Iowa City, Iowa 52242, USA ^dInterdisciplinary PhD Program in Genetics, Department of Otolaryngology, University of Iowa, Iowa City, IA 52242, USA ^eDepartment of Otorhinolaryngology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands ^fDepartment of Electrical and Computer Engineering, The University of Iowa, Iowa City, Iowa 52242, USA ^gDepartment of Ophthalmology and Visual Sciences, The University of Iowa, Iowa City, Iowa 52242, USA

Keywords

Audioprofiling; audiogram; ADNSHL; DFNA loci

Gene Prediction for Autosomal Dominant Non-syndromic Hearing Loss

Significant hearing loss (≥ 25 decibels) is prevalent in the adult population (15–20%) and affects approximately 50% of individuals 80 years of age or older¹. It is estimated that 15% of inherited hearing impairment is accounted for by autosomal dominant non-syndromic forms of hearing loss (ADNSHL). To date, 57 loci for autosomal dominant deafness have been mapped to chromosomal regions, and 23 causally-related genes have been identified at these loci (<http://webh01.ua.ac.be/hhh/>; May 8th, 2009)². In most cases the hearing loss is sensorineural (SNHL) and non-syndromic.

Phenotype-genotype relationships in ADNSHL have been extensively studied³. Based on these data, it is clear that some correlations are very robust, such as the low-frequency audioprofile associated with *WFS1*-related hearing loss (DFNA6/14/38)⁴, while other correlations are more difficult to define. Autosomal dominant high-frequency hearing loss, for example, can be caused by mutations in a large number of different genes including *KCNQ4* (DFNA2), *DFNA5* (DFNA5), *COCH* (DFNA9) and *POU4F3* (DFNA15). By analyzing additional audiometric data at each of these loci, however, it may be possible to cluster genes that cause high-frequency hearing loss into definable subclusters. To determine whether these genes fall into phenotypically defined subclusters, multiple regression studies

Corresponding author: Prof. Richard Smith, Department of Otolaryngology - Head and Neck Surgery, University of Iowa, 5270 CBRB Building, Iowa City, IA 52242, USA, Tel: +319 335 6501, richard-smith@uiowa.edu.

[#]These authors contributed equally to this work.

No authors involved in this study report a conflict of interest.

of threshold data can be conducted taking into account age and/or select frequencies. This type of audioprofile subclustering could potentially increase the efficiency of gene identification in small families segregating ADNSHL.

To address this challenge we have developed and reported a computer program – AudioGene – that uses a machine-learning approach to analyze audioprofiles as a method of prioritizing genes for mutation screening in small families segregating ADNSHL5. The purpose of this communication is to describe improvements to our software and to introduce a new website (<http://audiogene.eng.uiowa.edu/audiogram>) we have developed to allow physician access to the AudioGene v3.0 software.

Training and Validation of AudioGene

Earlier, we reported the development of AudioGene v2.0 based on audiograms from 360 persons with ADNSHL caused by mutations in *KCNQ4* (DFNA2), *DFNA5* (DFNA5), *WFS1* (DFNA6/14/38), *TECTA* (DFNA8/12), *COCH* (DFNA9) or *COL11A2* (DFNA13)5. We showed that AudioGene v2.0 could accurately predict the correct gene based on a classifier trained on two audiograms for each of the members of the training sets for DFNA8 and DFNA9 in over 90% of cases5. The AudioGene dataset has now been expanded to include a total of 16 DFNA loci – DFNA2, DFNA5, DFNA6/14/38, DFNA8/12, DFNA9, DFNA10, DFNA11, DFNA13, DFNA15, DFNA18, DFNA20/26, DFNA25, DFNA28, DFNA36, DFNA41 and DFNA44 – with the current version of AudioGene (v3.0) using 1979 audiograms curated from 713 persons with hearing loss at one of these DFNA loci (Figure 2C).

To be useful, we hypothesized that AudioGene should be more accurate at gene prediction than human experts. We demonstrated this fact in an experiment in which 27 experts and AudioGene were used to analyze the same set of audiograms. The average accuracy of the human experts was 55% as compared to 88% for AudioGene v2.05. We next validated AudioGene v2.0 as a research tool by analyzing 160 individuals from 77 unrelated American families with high-frequency ADNSHL of unknown etiology. Mutations were predicted in *KCNQ4* (DFNA2) or *COCH* (DFNA9) in 48 and 6 families, respectively, and identified in 3/48 and 1/6 families, for a positive predictive value ranging from 6.3% to 16.7%5,6.

A Web Interface for AudioGene

We believe that the current version of AudioGene (v3.0) represents a useful clinical tool and therefore we have constructed a web interface to allow health-care providers to access this resource (Figure 2). The site is accessible at <http://audiogene.eng.uiowa.edu/audiogram> and does not require registration or a password. It includes links to upload de-identified patient and audiogram data (Table 1) and a dialogue box for the physician to input an e-mail address (Figure 2A). Once data are uploaded, a series of graphical plots are generated to allow the user to validate the audiograms (Figure 2B). Once the uploaded audiograms are verified as correct, the default analysis will compare the input audiograms against data in AudioGene v3.0. The user can also select the DFNA loci to be included in the analysis, although a minimum of three loci must be selected (Figure 2C).

The audiogram analysis uses a computational clustering algorithm to identify the three most likely mutated genes for each patient and displays these results on the website in addition to sending an e-mail notification to the health-care provider's e-mail address (Figure 2D). During the computational run, the results page is updated every 15 seconds until the analysis is complete. Because patient information and predicted results are confidential, the input

file, e-mail address and results are not stored once the analysis is complete. The steps required to run AudioGene v3.0 are outlined in Figure 3.

A key consideration for the health-care provider using AudioGene is how to incorporate the predictive results into clinical practice. The decision to continue with genetic testing to verify AudioGene predictions rests with the individual or family involved and also depends upon the availability of genetic testing for a given locus. A guide to mutation screening of causative genes at all of the DFNA loci included in AudioGene is given in Table 2. The data in this table were collated from the literature using a recently published paper that comprehensively reviewed NSHL7.

Current Limitations and Future Plans for AudioGene Development

There are several limitations to AudioGene v3.0. For example, the ability to predict loci represented by only a few audiograms in the dataset (e.g. DFNA25, DFNA28) is limited and predictions for those loci for which the gene is currently unknown (DFNA18, DFNA41) are of limited clinical value since confirmatory genetic testing cannot be done. To improve AudioGene as a clinical tool, we have developed a secure database system called the Collaborative Phenotype System (CPS). This system will give research scientists and health-care providers the ability to submit audiograms to increase the phenotype-genotype database integral to audioprofile analysis. CPS is also used to collect data on eye, cancer and autism spectrum disorders from human subjects. In collaboration with Guy Van Camp and colleagues, we are also adding a separate cluster/classifier for age-related hearing impairment (ARHI; presbycusis) based on audiometric data from a cohort of 1,174 individuals aged 55–65 years⁸. This addition should improve discrimination between NSHL and ARHI and may extend the utility of AudioGene classification beyond genetic forms of deafness.

Summary

AudioGene v3.0 predicts the genotypic cause of hearing loss from audiometric data. We hope this resource will be valuable to clinicians in the care of families segregating ADNSHL and to research scientists who are focusing on the identification of novel ADNSHL loci. The repository of collected data represents the most comprehensive available and we hope to eventually include all reported DFNA families. As the size of this database increases, the predictive accuracy of AudioGene should also improve.

Acknowledgments

We would like to acknowledge helpful feedback on AudioGene from Drs John Renton, Richard Gurgel, Andrew Heaford and Paul Lenkowski. We thank Dr. Guy Van Camp (Department of Medical Genetics, University of Antwerp) for providing human audiometric data for ARHL. RJH Smith supported this project in part with National Institutes of Health (NIH)-NIDCD RO1 DCO3544. M. Hildebrand is supported by an Australian National Health and Medical Research Council (NHMRC) Postdoctoral Training Fellowship.

Funding: This project was supported by NIH-NIDCD grant (RO1 DCO3544) (RJHS)

REFERENCES

1. Morton NE. Genetic epidemiology of hearing impairment. *Ann N Y Acad Sci.* 1991; 630:16–31. [PubMed: 1952587]
2. Van Camp G, Smith RJ. Hereditary Hearing Loss Homepage. 2007
3. Huygen, P.; Pauw, R.; Cremers, C. Genes, Hearing and Deafness. From Molecular Biology to Clinical Practice - Audiometric profiles associated with genetic nonsyndromal hearing impairment:

- a review and phenotype analysis. Martini, M. Stephens, D. Read, A.P. London, UK: Informa Healthcare; 2007.
4. Cryns K, Pfister M, Pennings RJ, Bom SJ, Flothmann K, Caethoven G, Kremer H, Schatteman I, Koln KA, Toth T, et al. Mutations in the WFS1 gene that cause low-frequency sensorineural hearing loss are small non-inactivating mutations. *Hum Genet.* 2002; 110:389–394. Epub 2002 Apr 9. [PubMed: 12073007]
 5. Hildebrand MS, Tack D, McMordie SJ, DeLuca A, Hur IA, Nishimura C, Huygen P, Casavant TL, Smith RJ. Audioprofile-directed screening identifies novel mutations in KCNQ4 causing hearing loss at the DFNA2 locus. *Genet Med.* 2008 In press.
 6. Hildebrand MS, Tack D, Deluca A, Hur IA, Van Rybroek JM, McMordie SJ, Muilenburg A, Hoskinson DP, Van Camp G, Pensak ML, et al. Mutation in the COCH gene is associated with superior semicircular canal dehiscence. *Am J Med Genet A.* 2009; 149A:280–285. [PubMed: 19161137]
 7. Hilgert N, Smith RJ, Van Camp G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? *Mutat Res.* 2009; 681:189–196. Epub 2008 Aug 29. [PubMed: 18804553]
 8. Demeester K, van Wieringen A, Hendrickx JJ, Topsakal V, Fransen E, van Laer L, Van Camp G, Van de Heyning P. Audiometric shape and presbycusis. *Int J Audiol.* 2009; 48:222–232. [PubMed: 19363723]

Figure 1. Genetic screening strategy based on audioprofiling of small families segregating ADNSHL

The first step in the AudioGene workflow is curation of audiometric data from patients into an Excel file using Prism software. A Perl preprocessor then converts this information into a Weka-readable format that is used to predict the likely causative gene using a semi-supervised support vector machine (SVM).

A

The screenshot shows the AudioGene website interface. On the left is a dark sidebar with the University of Iowa logo and a navigation menu. The main content area has the AudioGene logo and header information. The page title is 'Step 1 of 3: Upload Audiogram'. It includes instructions on file format (2003 excel spreadsheet), a link to a template, a file upload field, and an email address field with a 'next' button.

Navigation
Average Audiogram
Description
Publications
Audiogram Analysis
Bibliography
Contact Us
Home

start searching... **GO**

increase text size
decrease text size
print this page
text-only version

AudioGene
Center for Bioinformatics and Computational Biology@Ulowa
Department of Otolaryngology@Ulowa

Audiogram Analysis home

Step 1 of 3: Upload Audiogram

File type:
Please download the template below and fill in your data. Please be sure to save the file as a 2003 excel spreadsheet with the xls extension, and not as a 2008 excel file format with the.xlsx extension. The patient data goes on the first sheet and the audiograms for the patients are placed on the second sheet.
[template.xls](#)

Audiogram file:

The file uploaded needs to be in the format described at the top of this page.

Email address: *

The email address that you want your results to be sent to.

next

B**Step 2 of 3: Verify Uploaded Data****next**

The results of your form submission:

C

Step 3 of 3: Dataset

Your data will now be processed once you hit the submit button. In about 4-5 minutes you should receive an email with the results of the analysis. Below is the list of loci that your dataset will be ran against. You can remove from the dataset by unchecking the box.

Dataset		
☐	number of patients	number of audiograms
☒ DFNA2	268	599
☒ DFNA9	116	348
☒ DFNA5	74	344
☒ DFNA6	67	270
☒ DFNA8/12	50	95
☒ DFNA11	42	94
☒ DFNA10	38	68
☒ DFNA13	38	64
☒ DFNA15	8	29
☒ DFNA20/26	4	27
☒ DFNA36	1	10
☒ DFNA25	3	9
☒ DFNA28	1	8
☒ DFNA41	1	6
☒ DFNA44	1	4
☒ DFNA18	1	4

D

Figure 2. AudioGene audiogram analysis website

The AudioGene website (<http://audiogene.eng.uiowa.edu/audiogram>) can be used for audiogram analysis and candidate gene prediction in three simple steps. (A) In step 1, audiometric information from a patient or family is uploaded by the physician using a template that is provided on the site. (B) In step 2, graphical plots of the audiometric data are generated for verification by the physician. (C) In step 3, the physician selects the available audiometric data in the AudioGene training set for ≥ 3 DFNA loci to run the analysis. (D) An example output showing candidate gene prediction results from the audiogram analysis that appear on the website in real time and are forwarded to the physician's e-mail address. Three predictions are made for each patient. It is important to remember that these results are only predictions to be used as a guide for possible genetic testing.

Figure 3. Simple steps to complete AudioGene audiogram analysis

This flow diagram indicates the steps to complete to analyze audiometric data. The results of this analysis represent only predictions and the patient and health-care provider must decide whether genetic testing should be pursued. * The recommended query is against data for all DFNA loci in the dataset.

Table 1

Template tables for the patient information form (top) and audiogram form (bottom). Example data entries are shown.

ID (No confidential info)	Date of Birth	Sex (M/F/Unknown)	Age of Onset (Years)	Number of affected family members	Data Source#	Known Mutation Locus	Gene	Citation (Pubmed_id)
Individual 13	12-May-1980	M	10	15	Primary	Unknown	Unknown	None

Patient (enter on patient form first)	Age (Years)	Ear* (L/R/Better)	dB of loss at:	125 Hz	250 Hz	500 Hz	1000 Hz	1500 Hz	2000 Hz	4000 Hz	6000 Hz	8000 Hz
Individual 13	56	R			35	50	50	50	50	60		70
Individual 13	54	L			15	50	60	60	60	65		70

* If there is a significant difference in threshold between the two ears, the better hearing ear should only be entered

The data source can be primary (from

Table 2

Reported mutations in 16 ADNSHL genes

Gene	Locus	Number of mutations worldwide	Number of mutations in Caucasians	Number of exons	Number of exons with mutations [#]	Number of amplifiers [†]	Function in hearing	OMIM number [‡]
WFS1	DFNA6/14	19	13	8	2 (5,8)	6	Ion homeostasis	606201
KCNQ4	DFNA2	12	8	14	6 (1,4,5,6,7,10)	6	Ion homeostasis	603537
COCH	DFNA9	12	8	12	3 (4,5,12)	3	Extracellular matrix protein	603196
TECTA	DFNA8/12	8	7	23	7 (3,10,14,16,17,18,20)	7	Extracellular matrix protein	602574
ACTG1	DFNA20/26	6	6	6	4 (3,4,5,6)	2	Hair bundle, cytoskeletal formation	102560
EYA4	DFNA10	6	4	21	5 (9,12,13)	4	Transcription factor	603550
MYO7A	DFNA11	5	4	48	5 (7,13,17,21,22)	5	Hair bundle, motor protein	276903
DFNA5	DFNA5	4	2	10	0*	1	Unknown function	608798
POU4F3	DFNA15	3	2	2	1 (2)	2	Transcription factor	602460
COL11A2	DFNA13	2	2	66	2 (31,42)	2	Extracellular matrix protein	120290
TMC1	DFNA36	2	2	24	1 (19)	1	Hair bundle	606705
VGLUT3	DFNA25	1	2	12	1 (5)	1	Signalling of hair cells and neurons	607557
CCDC50	DFNA44	1	1	12	1 (11)	1	Cytoskeletal formation	607453
TFCP2L3	DFNA28	1	1	16	1 (13)	1	Transcription factor	608641
Unknown	DFNA18	N/A	N/A	N/A	N/A	N/A	N/A	606012
Unknown	DFNA41	N/A	N/A	N/A	N/A	N/A	N/A	608224

[#]The exons containing reported mutations are shown in brackets

[†] DNA amplimers required to complete genetic screening for known mutations

[‡] Online Mammalian Inheritance in Man identification number

* All four deafness-causing mutations in DFNA5 are intronic and result in skipping of exon 8
Relevant references for these mutations have been summarized in ¹¹.