UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Assessing Gene-by-Smoking Interactions on Hearing Loss in Japanese Americans: A Candidate Gene Study

Permalink

https://escholarship.org/uc/item/1jg764nh

Author

Liem, Andrew

Publication Date

2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Assessing Gene-by-Smoking Interactions on Hearing Loss in Japanese Americans: A Candidate Gene Study

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Epidemiology

by

Andrew H. Liem

Thesis Committee:
Professor Karen Edwards, Chair
Assistant Professor Trina Norden-Krichmar
Assistant Professor Luohua Jiang

2017

© 2017 Andrew H. Liem

TABLE OF CONTENTS

	Page
LIST OF TABLES	iii
LIST OF FIGURES	iv
ACKNOWLEDGMENTS	v
ABSTRACT OF THE THESIS	vi
INTRODUCTION	1
METHODS	3
Study Population	
Phenotype: Outcome, Exposure, and Confounders	
Genotyping	
Selection of Single-Nucleotide Polymorphisms (SNPs)	
Statistical analysis	
RESULTS	8
DISCUSSION	10
TABLES	16
FIGURES	22
REFERENCES	25
APPENDIX A: Linkage evidence of gene-by-environment interaction in subset of families	30
APPENDIX B: Interpreting Logistic Models with Interactions	31

LIST OF TABLES

		Page
Table 1	Characteristic of Japanese American Families	16
Table 2	Characteristic of HLD (+) and HLD(-) members of Japanese-American families	17
Table 3	Variants under OSA Linkage Signals	18
Table 4	Genotype frequencies of markers for six candidate genes in the (+) and (–) HLD groups	19
Table 5	Beta coefficients from logistic regression analysis	20
Table 6	Odds ratio for Main and Interaction Effects	21

LIST OF FIGURES

		Page
Figure 1	Physiology of Hearing	22
Figure 2	Flow Chart of participants who met inclusion/exclusion criteria	23
Figure 3	Observing effect of GNAS variants in Smokers on Hearing Loss	24

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Karen Edwards, and my committee members, Assistant Professor Trina M. Norden-Krichmar and Assistant Professor Luohua Jiang. Their respective expertise in epidemiology, bioinformatics, and statistical analysis contributes to the success of this project. They have been knowledgeable and approachable and without their tutelage, this thesis would not have been possible.

I would like to thank members of Professor Karen Edwards research team. This includes, Emileigh Willems, Professor Stephanie A. Santorico at the University of Colorado, Denver, and most notably, Angel Wan. Participation in their weekly team meetings has provided guidance on my thesis project. I would also like to thank Christina Catabay, a previous University of California, Irvine graduate student that performed the preliminary analyses my hypothesis is based upon.

I would like to thank Professor and Chair Hoda Anton-Culver for developing an epidemiology program at the University of California, Irvine focused in genetics. Her dedication to each student in her program is commendable and exemplifies her commitment to education.

I would like to thank the families that participated in the American Diabetes Association sponsored study "Genetics of Non-insulin dependent diabetes mellitus (GENNID) study. I also wish to thank the American Diabetes Association for providing access to this data, and to the NIH; National Heart, Lung and Blood Institute for funding aspects of this work (R01HL113189; Edwards KL, PI). Without the families and the ADA, the analysis would not have been possible.

ABSTRACT OF THE THESIS

Assessing Gene-by-Smoking Interactions on Hearing Loss in Japanese Americans:
A Candidate Gene Study

By

Andrew H. Liem

Master of Science in Epidemiology

University of California, Irvine, 2017

Professor Karen Edwards, Chair

The prevalence of hearing loss and cigarette smoking poses major public health concerns. Studies have shown conflicting results between the association of smoking and hearing loss. As a complex, multifactorial disease, the discrepancy may be attributed to genetic modifiers that modify the relationship. Preliminary results using a genome-wide ordered subset analysis (OSA) analyzed Japanese American families and detected six regions on the genome showing evidence for gene-by-smoking interactions: 8q21, 7p21, 11q23, 12q32, 15q26, and 20q13. In this study, a candidate gene approach is used to specifically assess which genes may be driving the signals. A sample of 17 Japanese American families (N = 129) from the Genetics of Non-insulin Dependent Diabetes Mellitus (GENNID) study were selected and genotyped for analysis. Markers within, or close to 19 candidate genes previously implicated for hearing less were tested for interaction using logistic regression models. Generalized estimating equations adjusted for within-family correlations. Six genes were found to be significantly modifying the association between smoking and hearing loss: GNAS, RIPK2, GRLH2, RDX, EYA1, and HOMER2 with GNAS

significant after a Bonferroni correction. GNAS, GRHL2, and RDX significantly predicted hearing loss, independent of the effects of aging and smoking. This report provides evidence that the effect of smoking on hearing loss can be dependent on genotype. With a high prevalence of hearing loss and cigarette smoking worldwide and in the U.S, exploring this relationship can be useful for strategies to prevent hearing loss and reduce cigarette usage.

INTRODUCTION

Hearing loss and deafness (HLD) is a major public health concern. It is the most common sensory disorder worldwide and in the United States. Nearly 15% of the world's population (1+ billion) suffers some form of hearing loss[1]. In the U.S, the prevalence is estimated to be 20% of the population (48 million Americans)[2].

Epidemiological studies show the distribution of HLD is unequal in the U.S. HLD is inversely associated with education and income level. It is more prevalent in men than women and disproportionately affect the elderly[3–5]. The National Health and Nutritional Examination Surveys (NHANES) found that 60% of all Americans affected with HLD were 60 years of age or older making age-related hearing loss the most common form of HLD. This type of HLD, Presbyacusis, is likely caused by pathophysiological changes related to aging and cumulative damage acquired over the years from ototoxins and noise exposures. With a rapidly aging population, the prevalence of HLD is expected to rise in the following years.

There have been conflicting results about the association between HLD and other diseases and risk factors. While some studies report greater prevalence of HLD in groups with cardiovascular risk, diabetes, and smoking usage[4,6], others have not[5]. This could reflect HLD's complex, multifactorial etiology. Hearing impairment is driven by the effects of multiple genetic and non-genetic factors. Non-genetic factors include exposure to noise, aging, ototoxic drugs, and microbial infections[7,8]. However, the heritability of HLD (h^2 = 0.36) estimates that about a third of the disease is attributed to genetics[9]. The Hereditary Hearing Loss Homepage (http://hereditaryhearingloss.org) have collected more than a hundred loci and genes localized and identified for hearing loss.

Previous studies have found that cigarette smoking is associated with hearing loss and deafness[6,10–12]. Few biological mechanisms have been proposed to explain this link. Cigarette smoke contains over 7000 compounds, some of which are toxic to organs of the auditory system, such as hydrogen cyanide, carbon monoxide, and nicotine. These toxins can obstruct blood flow and directly damage hair cells of the cochlea. The cochlea is the sense organ of the inner ear that translates sounds into nerve impulses transmitted to the brain. Hair cells are the sensory receptors of the cochlea that pick-up vibrations from sound waves to generate the nerve impulses. Directly damaging hair cells diminishes its sensitivity to detect sound waves resulting in sensorineural hearing loss, the most common cause of HLD (Fig. 1)[12]. Indirectly damaging the cochlea can also lead to hearing loss. The cochlea is highly metabolic and lacks collateral vessels. Loss of function can occur when blood flow is obstructed[13]. Chronic smoking induces hemolysis, vasoconstriction, and increases carboxyhemoglobin levels and blood viscosity, which can lead to ischemia and hypoxia.

In a previous study, the joint effect of genetic and smoking on HLD was investigated using linkage analyses of Japanese American families. We looked at the effect of cigarette smoking, the exposure variable, on risk of HLD, the outcome variable, and whether the association is modified by human genetics. An ordered-subset analysis (OSA) ranked multiplex families (≥2 HLD cases) by average packs-per-day smoked or percent of ever smokers and found suggestive evidence for a gene-by-smoking interaction. The OSA approach recognizes and adjust for genetic and etiologic heterogeneity observed in complex, multifactorial diseases[14]. It aims to create groups genetically more homogenous using a function of a covariate to rank and subset families. This can reduce bias and

increase evidence of linkage. OSA found evidence of linkage on six loci: 8q21 (LOD=3.54), 7p21 (LOD=3.3), 11q23 (LOD=2.1), 12q32 (LOD=2.6), 15q26 (LOD=2.9), 20q13 (LOD=3) (see APPENDIX A). The region on chromosome 8 was identified using families with the highest average cigarette packs-per-day values. The regions on chromosome 7, 11, 12, 15 and 20 were identified using families with highest per-family percentage of smoking.

To determine what may be driving the linkage signals, a more comprehensive assessment is needed. We hypothesized that candidate genes in the linkage regions are driving the signals and interacting with smoking to modify risk for HLD. In this study, logistic regression models with generalized estimating equations assessed whether the association between smoking and hearing loss is dependent on genotype.

Understanding how genetics and smoking may interact to modify the risk of hearing loss can have huge public health impact. Like HLD, smoking is highly prevalent and the #1 cause of preventable mortalities in the U.S[15]. The incidence of cigarette usage is greater in males, diabetics, the elderly population, and those with lesser education[16–18].

METHODS

Study Population:

The Genetics of Non-Insulin Diabetes Mellitus (GENNID) project[19] was established in 1993 by the American Diabetes Association to study the genetic basis of diabetes. In its totality, it contains comprehensive information on more than 220 families (N=~1,400). The study is a diverse cohort of people with Non-Hispanic White, Hispanic, African American, and Japanese American origins. The preliminary study focused on Japanese

Americans because this subgroup contained the greatest number of HLD cases. GENNID contained a total of 17 Japanese American families with 15 families having at least one case of HLD and eight families having at least two. To increase power for detecting gene-by-environment interaction by linkage, the eight multiplex families (N = 97) were used for the linkage analysis. For the association test, all 17 Japanese American families (N = 129) were used to maximize the sample size.

Phenotypes: Outcome, Exposure and Confounders:

Data for each subject was collected from self-reported questionnaires and laboratory tests. This provided information on medical and family histories, diabetes status, education level, and smoking usage. In GENNID, HLD status was a binary trait (yes or no). It was evaluated with the question "Have you been told by a health care worker that you have hearing loss/deafness?" Smoking status was also a binary trait evaluated by "Do you now or have you ever used any form of tobacco?" If answered 'Yes,' participants were asked what forms of tobacco were used: cigarettes, cigars, pipe, or chew/snuff. The questionnaire requested per type age when they began, age stopped, current usage status, and the average amount used per day. Dose of usage for cigarettes, cigars, pipe, and chew/snuff were measured in units of packs/day, number/day, bowls/day, and wads/day, respectively.

Four covariates were identified as confounders and were tested in the model: age, sex, diabetes, and education level. In this study, a confounder is defined per Rothman and Greenland as a variable associated with the exposure and outcome variable, but not laying on the causal pathway. Diabetes status was evaluated with the following question "Are you now or have you been told by a medical professional that you are diabetic?" Sex was either

male or female. Education level was collected using five levels: less then High School graduate, High School graduate, technical school graduate, college graduate, or advanced degree.

Genotyping

In the previous analysis, 374 polymorphic microsatellite markers were used to identify the six linkage peaks. Quality control processes were performed using PEDCHECK to assess genotype errors and Mendelian inconsistencies in family pedigrees[20]. For this study, single nucleotide polymorphisms (SNPs) level information provided by Illumina's Infinium® Multi-Ethnic Global BeadChip were used. This array provided information on 1.7 million SNPs. Quality control was completed using standard approaches as part of a larger project. This include first removing anomalous segments and samples based on sex discrepancy and quality scores, removing contaminated samples with high heterozygosity (> mean + 4SD), and removing samples and variants with >2% missing. Duplicate samples, batch effect, and Mendelian errors were also checked. All monomorphic markers were removed. Variants were also removed if they failed to meet Hardy-Weinberg equilibrium, had more than two Mendelian errors, and had a minor allele frequency <1%.

Selection of Single-Nucleotide Polymorphism (SNPs)

One-LOD support intervals were constructed for each linkage regions. The linkage regions on chromosome 8 and 11 covered more than 40 million nucleotide base pairs. Chromosome 7, 15, and 20 contained at least 15 million base pairs with chromosome 12 having the smallest range of 2.6 million nucleotides (Table 4). Number of SNPs in the linkage regions were provided by the Single Nucleotide Polymorphism database[21] (dbSNP

Build 149), a public-archive managed by the National Institute of Health. A total of eight million SNPs and 5000 genes were found in the six linkage regions.

To reduce the number of genes to test, a candidate gene approach was used. This hypothesis-driven approach is more powerful than an untargeted screening strategy. A candidate gene was defined as any protein-coding, gene previously implicated for HLD. SNPs considered related to a gene were either within the gene sequence or 1MB of the gene ends. This range is conventionally used to define *cis* expression quantitative trait loci (cis-eQTL)[22]. Only SNPs within this range were tested for gene-environment interaction. We did not narrow down to specific functional elements of the gene for testing because we believed not all functional elements have been completely mapped. The causative polymorphism may be lying in a 'non-functional' region. An unadjusted P-value < 0.05 was considered as significant because a candidate gene approach was used and we expect the effect size to be relatively small for gene-environment interactions. A general rule of thumb recommends a sample size at least four times larger to detect an interaction effect of the same magnitude as the main effect[23]. Variants were prioritized by P-value and adjusted for multiple testing by a Bonferroni correction.

Genes previously implicated for hearing loss were identified in two ways: reviewing the literature and databases. A literature review on PUBMED and Google Scholar was conducted using key words: deafness genes, hearing loss genes, genes associated with hearing loss, and genes associated with deafness. Genotype-phenotype databases including Phenopedia[24], the Online Mendelian Inheritance in Man (OMIM) [25], and a specialized data base (Hereditary Hearing Loss Homepage) were scanned.

In total, 502 HLD genes supported by at least one publication were found across the genome. Of these 502 genes, 92 were on the same chromosome as the OSA linkage signals and 19 were found within the one-LOD support interval regions (Table 4). No candidate genes were found in the one-LOD support interval regions on chromosome 7 or 12. Markers were pruned by linkage disequilibrium ($r^2 > 0.8$) and manually selected using the HapMap-JPT population on the Genome Variation Server. Decision on which variant to analyze per bin was determined using the VarSome database. Variants considered most likely to be pathogenic were selected for and evaluated by computational tools (e.g. DANN, Protean), conservation score, allele frequency, and functional studies. Before LD pruning, a total of 310 variants passed quality control. Pruning by linkage disequilibrium removed 49 variants leaving 261 variants for testing.

Statistical analysis

Statistical analyses of eight Japanese American families assessed the interaction between genetic variants and smoking on HLD risk. A multivariable logistic regression model was used because the dependent variable is a dichotomous trait. Generalized estimating equations (GEE) with an exchangeable correlation structure adjusted for within-family correlations in the data. This structure is best fit for family studies than other structures (autoregressive, unstructured, independent)[26] and have been used previously for association tests using related individuals[27–30].

Our study is interested in β_3, the interaction coefficient, and whether it is statistical significant by Wald's Z-test at $\alpha = 0.05$. The null hypothesis states there is no difference in the effect of genotype among smokers and nonsmokers on HLD risk. If significant, the effect of smoking on HLD is dependent on genotype. We chose an association test at the

genotypic and not allelic level. An allelic test can provide twice the sample size and has fewer degrees of freedom, but it is an observation at the gamete and not the individual. Its validity is also compromised if alleles are not in Hardy-Weinberg Equilibrium. We related the risk of HLD given an individual's genotype with an additive model. The smoking trait in the analysis is a continuous variable (units in cigarette packs-per-day). We identified four variables as potential confounders: age of participant at diagnosis (years), sex (1= male, 0 = female), diabetes diagnosis (1 = yes, 0 = no), and education status (0 = less than high school graduate, 1 = high school graduate, 2 = technical school graduate, 3 = college graduate, 4 = advanced degree). Diabetes status was self-reported by the following question, "Are you now or have you been told by a medical professional that you are diabetic?" In total, we started with four categorical variables (one ordinal), two continuous variables, and one interaction term to build our model. All analyses were performed with R software version 3.3.2 using the geepack package. Interaction was conceptualized in terms of odds ratios. See appendix for details in interpreting a logistic regression model.

Results

GENNID contained 17 Japanese-American families. There was a median of 7 subjects per family and a total of 129 family members (Table 1). In our data, univariate analysis found that subjects with HLD were more likely to be older ($P = 1.3 \times 10^{-5}$), male (P = 0.01), and have been ever smokers (P = 0.03) (Table 2). Odds of HLD for ever smokers were 2.31 (P-value = 7.3×10^{-3}) relative to odds of never smokers. In our data, we did not find a significant difference in education level or diabetes status between participants although when adjusted for age, education level was significant (P = 0.038). In building our

model, we put cigarette smoking first because it is our exposure of interest. Variables for the model were then selected by manual section based on statistical significance. These variables were potential confounders: age, diabetes, education, and sex. Adding education level and diabetes status individually were insignificant while age ($P = 8.8 \times 10^{-4}$) and sex ($P = 0.076$) were significant. Age was added to the model instead of sex because it had a smaller P-value. Adding sex with age was insignificant ($P = 0.129$).

The final fitted model adjusted for age and is described as followed:

$$\text{Logit}\left(\text{Probability of HLD}_{ij}\right) =$$

$$\beta_0 + \beta_1\left(\text{Smoking}_{ij}\right) + \beta_2\left(\text{Genotype}_{ij}\right) + \beta_3\left(\text{Smoking}_{ij} * \text{Genotype}_{ij}\right) + \beta_4(\text{Age}_{ij})$$

In the expression, *i* refers to family unit and *j* refers to *j*th person from the *i*th family.

A separate regression analysis was performed for each marker. Six genes were significant for an interaction with an unadjusted $\alpha = 0.05$: EYA1 (P-value = 0.03), RIPK2 (P-value = 1.0×10^{-3}), GRLH2 (P-value = 8.0×10^{-3}), RDX (P-value = 0.03), HOMER2 (P-value = 0.04), and GNAS (P-value = 1.80×10^{-16}). The genotype frequencies of these markers stratified by HLD groups are provided in Table 4. It is interesting to note that half of these genes were on chromosome 8, the most significant of the OSA linkage regions (LOD 3.54). Table 5 provides a summary of the top SNPs by P-value for interaction per gene expressed in beta coefficients. Only GNAS (guanine nucleotide binding protein-α subunit) was significant after Bonferroni correction ($P < 1.9 \times 10^{-4}$). The most significant marker for

GNAS was on position 57578447 (hg19 build) with adenine as the minor allele. Results show that this minor allele in GNAS has a protective effect. For every minor allele a smoker has, their odds for HLD is reduced by 0.355 under an additive model. Smokers without the genotype have an estimated odds ratio of 22.9 (P-value = 2.1×10^{-5}). The top scoring marker for gene RDX (rs6988306) also had a protective effect with OR = 0.288 in smokers with the variant versus OR = 15.5 (P-value = 0.014) for smokers without the risk variant. The most significant markers for EYA1 (rs10957540), RIPK2 (rs39765), GRHL2 (rs6988306), and HOMER2 (15:830509561) increased the risk of HLD in ever smokers. Both main genotype and smoking effect in the models testing for GNAS and RDX were significant, but not for EYA1, GRHL2, HOMER2, RIPK2. For EYA1, HOMER2, and RIPK2, the main effect of the genotype and smoking variable were insignificant. This suggests the effect of the genes are conditional on whether an individual has ever smoked. For GRHL2 in which the genotype main effect was significant and the smoking main effect was not, the gene has an effect even when the individual has never smoked. Overall, a total of six candidate genes were significant with an unadjusted $\alpha = 0.05$ with one of the genes, GNAS, significant with a Bonferroni correction

Discussion

In this study, a linkage analysis using polymorphic markers informed which genomic regions to investigate for closer inspection. We applied logistic regression models using generalized estimating equations to assess if candidate genes were driving the gene-environment interaction signals. GNAS, RIPK2, GRHL2, RDX, EYA1, and HOMER2 were significant for an interaction with GNAS remaining significant after a Bonferroni correction

(GNAS). Overall, the candidate gene approach evaluated four of the six OSA regions and found genes in three of the four regions investigated. These results support evidence that the association between smoking and HLD can be dependent on genotype. Possible future investigations would be performing an untargeted screening strategy to evaluate what may be driving the signals in chromosome 7 and 12 and what else may be driving the linkage signals in chromosome 8, 11, 15, and 20. It is possible that our candidate gene explains only proportion of the linkage signal. In an untargeted screening strategy, these significant genes could be conditioned upon to evaluate what other genes may be driving the signal. Broadening association tests to noncoding regions would be also useful as dysfunction in the regulatory network plays a role in etiology.

Of the significant genes, GNAS and RDX are noteworthy. These genes were significant for an interaction and situated closest to the max LOD marker within their respective chromosome. GNAS, the most significant hit, lies closest to the 20q13 marker (LOD = 3.0) 0.3Mb downstream. The variant lies in an intronic region at the upstream end of the gene near its promoter (Figure 3). The variant may be altering expression of GNAS. Evaluation of other variants close or within GNAS also were disproportionately significant ($\alpha = 0.05$) at the upstream end and showing a protective effect (OR < 1). In the literature, mutations in the gene are known to cause thyroid conditions such as pseudohypoparathyroidism[31]. The gene is strongly associated ($P = 1.3x10^{-12}$) with levels of thyroid-stimulating hormones (TSH), a critical regulator of hormone production in the thyroid gland[32]. A strong link has been found between hearing loss and imbalanced thyroid levels. Children with hypothyroidism experience greater hearing loss than the general population[33] and mouse models have shown insufficient levels of thyroid hormones can

block hearing development[34,35]. Low TSH levels are also characteristics of current smokers[36,37]. Compounds of cigarette smoking like nicotine affects the endocrine system leading to development of various diseases. Smokers are more likely to have enlarged thyroids, an indication of abnormal thyroid levels. It may be that Japanese American families in GENNID possess a GNAS variant (20: 57478447) protective against the effect of smoking on TSH level enabling normal development and function of hearing. If GNAS is part of one pathway that explains the link between smoking and HLD, an avenue to care for effected individuals are pharmacological treatments typically for thyroid dysfunction. Interestingly, administration of levothyroxine have been observed to improve and sometimes reverse congenital hearing loss[38]. In addition to GNAS, RDX also sits close to the max LOD marker at 1.1Gb downstream from the 11q23 marker (LOD = 2.1). RDX codes for a cytoskeletal protein expressed in the human retina and inner ear. It is localized along the inner ear hair cell stereocilia and plays a critical role in anchoring membrane proteins to actin filaments[39]. In mouse models, knocking out RDX resulted in deafness[40,41]. The damaging effect of cigarette smoking on inner ear hair cells may be exacerbated if its structural integrity is compromised. In our study, RDX had a significant interaction with smoking (P-value = 0.030) and a significant main effect (P-value = 0.024). The marker rs7932174 for RDX increased the risk of hearing loss in ever smokers than never smokers (OR = 2.14). This polymorphism may affect the protein derivative of RDX reducing its ability to withstand smoking.

This study benefits from several strengths including focusing on targeted areas in a joint linkage and association analysis approach, using family units, and concentrating only on Japanese Americans individuals. Association studies have high power to detect common

variants with small effect for a phenotype while linkage studies have high power to detect regions containing polymorphisms of large effect. Combining these techniques can help detect variants with the largest effect size and common across families. This also helps narrow down targeted areas for investigation which provides a hypothesis-driven approach than an untargeted scan. This increases power and reduces type I errors. Moreover, we believe we have increased power using families than unrelated individuals by reducing heterogeneity. There are different subtypes of HLD which could be inherited as an autosomal recessive or dominant model, X-linked, or of mitochondrial origin[42]. Each subtype could reflect heterogeneity in the pathogenesis of hearing loss or a different underlying genetic architecture. Using families for the analysis likely reduces etiologic heterogeneity caused by different subtypes of HLD. Focusing only on Japanese Americans also help reduce heterogeneity caused by ethnic-related differences.

Despite our strengths, there are a number of limitations. There is reduced power from misclassification and biases due to self-reporting and we are confronted with the general challenges of detecting gene-environment interactions. HLD status was obtained by self-report questionnaires as either 'yes' or 'no' for hearing loss. This method was cost-effective and easier for data collection than bringing a trained audiologist, but response bias may occur if the participant felt pressured to answer inaccurately. They could be basing their answer on what is socially acceptable or what they perceive to be reality, but is not. In the questionnaire, participants may have underreported smoking usage because of its known health risk resulting in misclassification. There is also loss of power using binary and categorical variables. Audiometry test administered by an audiologist could have better classified the extent of hearing loss to improve power using a continuous variable

and reduce misclassification error. Hearing loss can be measured in decibels (dB) or by percent of hearing impairment measured in Hertz (Hz). The American Speech-Language Hearing Association (ASHA) for an example classify hearing loss on a spectrum from normal to severe according to decibel ranges: normal (-10 - 15 dB), slight (16 – 25 dB), mild (26 – 40 dB), moderate (41 – 55 dB), moderately severe (56 – 70 dB), severe (71 – 90 dB), and profound (91+)[43]. Otologic, ancillary, and molecular genetic testing have also evaluated severity and subtype of HLD.

This gene-environment interaction shares similar challenges to other gene-environment studies. Having a large sample size is a challenge. The data provided 29 cases of HLD and there were many more individuals with non-risk genotypes than with risk genotypes (i.e. heterozygotes and rare homozygotes). Some markers did not have any individuals homozygous for the minor allele. There were overall low counts of risk genotypes for all markers of interest. While this was expected, it does reduce the power from small-sample bias. Lastly, there exist challenges in measuring gene activity and environmental exposure simultaneously. Gene expression and environmental exposure both varies by time and place and an interaction requires concurrent activity. It is difficult to measure the frequency and intensity of an environmental exposure concurrently with gene expression levels.

Overall, this study considered the role of genetics in the effect of smoking on HLD and found evidence of its modifying effect. Focusing on Japanese American and using related individuals helped reduce heterogeneity to increase power. GNAS was the most significant gene followed by RIPK2, GRLH2, RDX, EYA1, and HOMER2. GNAS's role in TSH suggest treatments for thyroid conditions maybe one avenue for treatment of smoking-

induced HLD. Our work requires replication in a separate cohort, but it supports the need to consider gene-environment interaction to explain the risk of HLD. Considering the highly prevalent nature of cigarette smoking and HLD, identifying genetic modifiers has the potential to have a large public health impact.

Tables

Table 1: Characteristic of 17 Japanese-American Families from GENNID (Genetics of Non-Insulin Dependent Diabetes Mellitus) study, sponsored by the American Diabetes Association. Shown are number of family members with diabetes, deafness, and has ever smoked.

Family ID	# of family members	Avg. age (SD)	+diabetes	+HLD	ever smokers
Family 1	15	49.8 (8.8)	2	2	5
Family 2	14	53.5 (11.3)	4	4	11
Family 3	11	51.3 (16.3)	1	1	7
Family 4	11	47.6 (19.8)	3	4	2
Family 5	9	59.9 (15.1)	1	2	7
Family 6	8	53.1 (16.1)	2	1	4
Family 7	8	58.2 (16.8)	4	2	5
Family 8	7	49.6 (16.8)	2	1	4
Family 9	7	66.8 (8.3)	4	4	6
Family 10	7	60.4 (13.6)	3	1	3
Family 11	7	52.2 (15.3)	1	2	4
Family 12	6	62.95 (13.5)	2	2	5
Family 13	6	68.9 (14.1)	2	0	1
Family 14	6	69.7 (9.1)	4	0	2
Family 15	4	62.9 (16.8)	1	2	1
Family 16	2	68.1 (8.8)	0	1	0
Family 17	1	66.0 (N/A)	0	0	1
Total	129	56.51 (15.7)	36	29	68

Table 2: Characteristic of HLD(+) and HLD(-) members of Japanese-American families

GEE logistic regression to test differences between HLD+ and HLD- subgroups. Two models were tested. Model 2 adjusted for age. Insulin and glucose was measured in mcU/mL and mg/dL, respectively.
* notes significant <0.05

		HLD+ (N=29)	HLD- (N=100)	All subjects (N=129)	P (unadj.)	P (age adj.)
	Age (± SD)	66.45 ± 9.98	53.63 ± 15.97	56.51 ± 15.74	1.3×10^{-5}*	N/A
	Sex (%), **# of males**	20 (68%)	47 (47%)	67 (52%)	1.3×10^{-2}*	0.057
	Age of Deafness (± SD)	52.2 ± 18.5	—	—	N/A	N/A
Diabetes characteristics	**Diabetes** (%)	12 (41%)	24 (24%)	36 (27%)	0.11	0.85
	Avg. Insulin (± SD)	8.2 ± 4.9	7.2 ± 6.7	7.4 ± 6.3	0.48	0.89
	Avg. Glucose (± SD)	126.3 ± 45.6	113.3 ± 27.8	116 ± 33	0.26	0.48
Smoking characteristics	**Cigarette ever smoker** (%)	21 (72%)	47 (47%)	68 (52%)	2.7×10^{-2}*	0.064
	Age began smoking (%)	18.6 ± 4.5	20.5 ± 6.9	19.9 ± 6.4	0.25	0.084
	Age stopped smoking (%)	42.5 ± 11.4	41.3 ± 14.2	41.6 ± 13.4	0.56	0.577
	Packs-per-day	0.75 ± 0.79	0.38 ± 0.56	0.46 ± 0.63	7.3×10^{-3}	0.088
Education (%), **(college graduate+)**		15 (51%)	52 (52%)	67 (52%)	0.74	0.048*

Table 3: Variants under OSA Linkage Signals:

Number of variants in each linkage peak obtained from dbSNP build 147 (Feb. 2009, GRCh37/hg19). There are 4 sets of number representing four different SNP tracks. (1) 'all SNPs' refer to *all* mappings of reference SNPs to the human assembly. (2) 'Common SNPs' contains all variants that appear in at least 1% of the population or are 100% non-reference (3) 'Flagged SNPs' contains all *flagged* SNPs that are clinically associated (excludes common SNPs). (4) 'Genes' obtained from Ensembl Genes track

CHR	OSA max LOD Marker (1-LOD flanking markers)	Base pair location (1-LOD flanking BP)	All SNPs	Common SNPs	Flagged SNPs	Genes
8	D8S1119 (D8S1136 - D8S1132)	87172020 (66,053,167- 107,328,872)	2,067,933	189,252	1,228	2,081
7	D7S3051 (D7S1819 - D7S1802)	18284528 (4,493,039 - 20,705,304)	1,001,982	106,992	682	281
11	D11S1986 (D11S2002 - D11S4464)	111223605 (79,965,403 -123,626,627)	2,335,240	222,100	2,784	975
12	D12S1045 (D12S2078 - D12S392)	130397738 (127,961,140 - 130,607,434)	156,279	18,911	70	19
15	D15S16 (D15S1015 - D15S657)	95019784 (67,839,381 - 96,704,848)	1,559,130	138,491	917	1,152
20	D20S171 (D20S197 - rs608901)	57808030 (46,160,764 - 62,911,391)	988,398	93,515	655	540
Total			8,108,962	769,261	6,336	5,048

Table 4: Genotype frequencies of top markers for six candidate genes in the (+) and (–) HLD groups

Describes distribution of genotype across HLD groups. In the genotype column, 1 = minor allele and 0 = common allele. P-value of the genotype main effect from the final fitted model is shown.

Gene	Marker	Genotype	HLD groups						P-value (adj.)
			–	(%)	+	(%)	Total		
GNAS	(20:57578447)	1,1	2	(2.0)	1	(3.4)	3	(2.3)	
		1,0	22	(22)	5	(17.2)	27	(20.9)	
		0,0	76	(76)	23	(79.3)	99	(76.7)	0.012*
RIPK2	rs3976*5	1,1	0	(0.0)	0	(0.0)	0	(0.0)	
		1,0	4	(4.0)	3	(10.3)	7	(5.4)	
		0,0	96	(96)	26	(89.7)	122	(94.6)	0.208
GRLH2	rs529132	1,1	0	(0.0)	0	(0.0)	0	(0.0)	
		1,0	9	(9.0)	3	(10.3)	12	(9.3)	
		0,0	91	(91)	26	(89.7)	117	(90.7)	8.00×10^{-4}*
RDX	rs7932174	1,1	1	(1.0)	1	(3.4)	2	(1.6)	
		1,0	8	(8.0)	2	(6.9)	10	(7.8)	
		0,0	91	(91)	26	(89.7)	117	(90.7)	0.024*
EYA1	rs10957540	1,1	0	(0.0)	0	(0.0)	0	(0.0)	
		1,0	22	(22)	8	(27.6)	30	(23.3)	
		0,0	78	(78)	21	(72.4)	99	(76.7)	0.654
HOMER2	(15:830509561)	1,1	0	(0.0)	0	(0.0)	0	(0.0)	
		1,0	11	(11)	7	(24.1)	18	(14)	
		0,0	89	(89)	22	(75.9)	111	(86)	0.658

Table 5: Beta coefficients from logistic regression analyses with GEE.

Beta coefficients of smoking, genotype, and interaction. Only top SNPs for interaction (defined by lowest p-value) are reported per gene.
* denotes significance for interaction with P <0.05 (unadjusted)

	Marker	Consequence	Chr.	Gene	Interaction		Smoking ME		Genotype ME	
					β	P-value	*β*	P-value	*β*	P-value
1.	rs10957540*	intron	8	EYA1	1.806	0.034	-1.439	0.172	-0.341	0.654
2.	rs39765*	Downstream gene variant	8	RIPK2	1.149	0.001	-0.555	0.441	0.436	0.208
3.	rs529132*	Intron	8	GRHL2	2.939	0.008	-2.447	0.060	-1.993	0.008
4.	rs28928592	Intron	8	RRM2B	0.792	0.275	-0.218	0.808	-0.018	0.976
5.	11:92718649	Intron	11	MTNR1B	-1.122	0.345	1.831	0.148	2.072	8.48×10^{-6}
6.	rs76087982	Intron	11	MTMR2	1.742	0.106	-1.468	0.224	-0.873	0.302
7.	11:102668702	Intron	11	MMP1	0.055	0.934	0.554	0.675	-0.199	0.695
8.	rs77208579	Intron	11	CASP1	1.434	0.137	-0.904	0.398	-1.134	0.318
9.	rs7932174*	Intron	11	RDX	-1.949	0.030	2.744	0.014	0.704	0.024
10.	rs667125	Intron	11	TECTA	-0.747	0.159	1.768	0.079	0.291	0.514
11.	15:75015305	Intron	15	CYP1A1	0.348	0.552	0.190	0.849	-0.688	0.198
12.	15:78400673	Splice region, low	15	CIB2	0.501	0.344	0.043	0.950	0.422	0.351
13.	15:83509561*	Intron	15	HOMER2	1.727	0.043	-1.274	0.218	0.286	0.658
14.	rs4932187	Intron	15	ISG20	0.928	0.183	-0.731	0.552	-0.544	0.273
15.	rs56303756	Upstream gene variant	15	POLG	-1.075	0.073	1.825	0.065	0.180	0.869
16.	20: 57478447*	Intron	20	GNAS	-2.083	1.80×10^{-6}	3.129	2.1×10^{-5}	1.049	0.012
17.	rs3787432	Intron	20	OSBPL2	-0.939	0.161	2.386	0.124	0.867	0.069
18.	20: 61468063	Intron	20	COL9A3	-0.778	0.185	1.786	0.063	0.281	0.560
19.	N/A	N/A	20	PRPF6	N/A	N/A	N/A	N/A	N/A	N/A

Table 6: Odds ratio for Main and Interaction effects.

Odds ratio and significance level of the smoking, genotype (adjusted and unadjusted), and genotype-by-environment interaction. Only top SNPs for interaction (defined by lowest p-value) are reported per gene. ORs for genotype-by-smoking, smoking, and genotype effect were obtained by exponentiating combined value of β_1 and β_3, β_1, and β_2, respectively. (ME = main effect)

	Marker	Chr	Gene	Genotype ME (unadj.)		Genotype ME (adj.)		OR Smoking, Genotype [+]		OR Smoking, Genotype [–]	
				OR	P-value	OR	P-value	OR	P-value	OR	P-value
1.	rs10957540*	8	EYA1	2.217	0.518	0.711	0.654	4.328	0.034	0.237	0.172
2.	rs39765*	8	RIPK2	2.699	1.89×10^{-5}	1.547	0.208	4.880	0.001	0.574	0.441
3.	rs529132*	8	GRHL2	2.828	0.883	0.136	0.008	2.574	0.008	0.087	0.060
4.	rs28928592	8	RRM2B	1.923	0.447	0.982	0.976	2.169	0.275	0.804	0.808
5.	11:92718649	11	MTNR1B	3.196	0.028	7.941	8.48×10^{-6}	2.587	0.345	6.239	0.148
6.	rs76087982	11	MTMR2	0.875	0.288	0.418	0.302	2.383	0.106	0.230	0.224
7.	11:102668702	11	MMP1	2.204	0.667	0.820	0.695	0.866	0.934	1.741	0.675
8.	rs77208579	11	CASP1	1.228	0.630	0.322	0.318	1.349	0.137	0.405	0.398
9.	rs7932174*	11	RDX	2.139	0.509	2.023	0.024	0.288	0.030	15.552	0.014
10.	rs667125	11	TECTA	1.745	0.691	1.338	0.514	0.634	0.159	5.862	0.079
11.	15:75015305	15	CYP1A1	2.442	0.812	0.503	0.198	0.712	0.552	1.210	0.849
12.	15:78400673	15	CIB2	1.963	0.126	1.525	0.351	2.517	0.344	1.044	0.950
13.	15:83509561*	15	HOMER2	2.713	0.048	1.330	0.658	7.479	0.043	0.280	0.218
14.	rs4932187	15	ISG20	2.196	0.515	0.581	0.273	1.468	0.183	0.482	0.552
15.	rs56303756	15	POLG	1.679	0.777	1.197	0.869	0.408	0.073	6.202	0.065
16.	20: 57478447*	20	GNAS	2.020	0.788	2.854	0.012	0.355	1.80×10^{-6}	22.858	2.1×10^{-5}
17.	rs3787432	20	OSBPL2	1.575	0.021	2.380	0.069	0.930	0.161	10.872	0.124
18.	20: 61468063	20	COL9A3	1.754	0.715	1.324	0.560	0.608	0.185	5.967	0.063
19.	N/A	20	PRPF6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figures

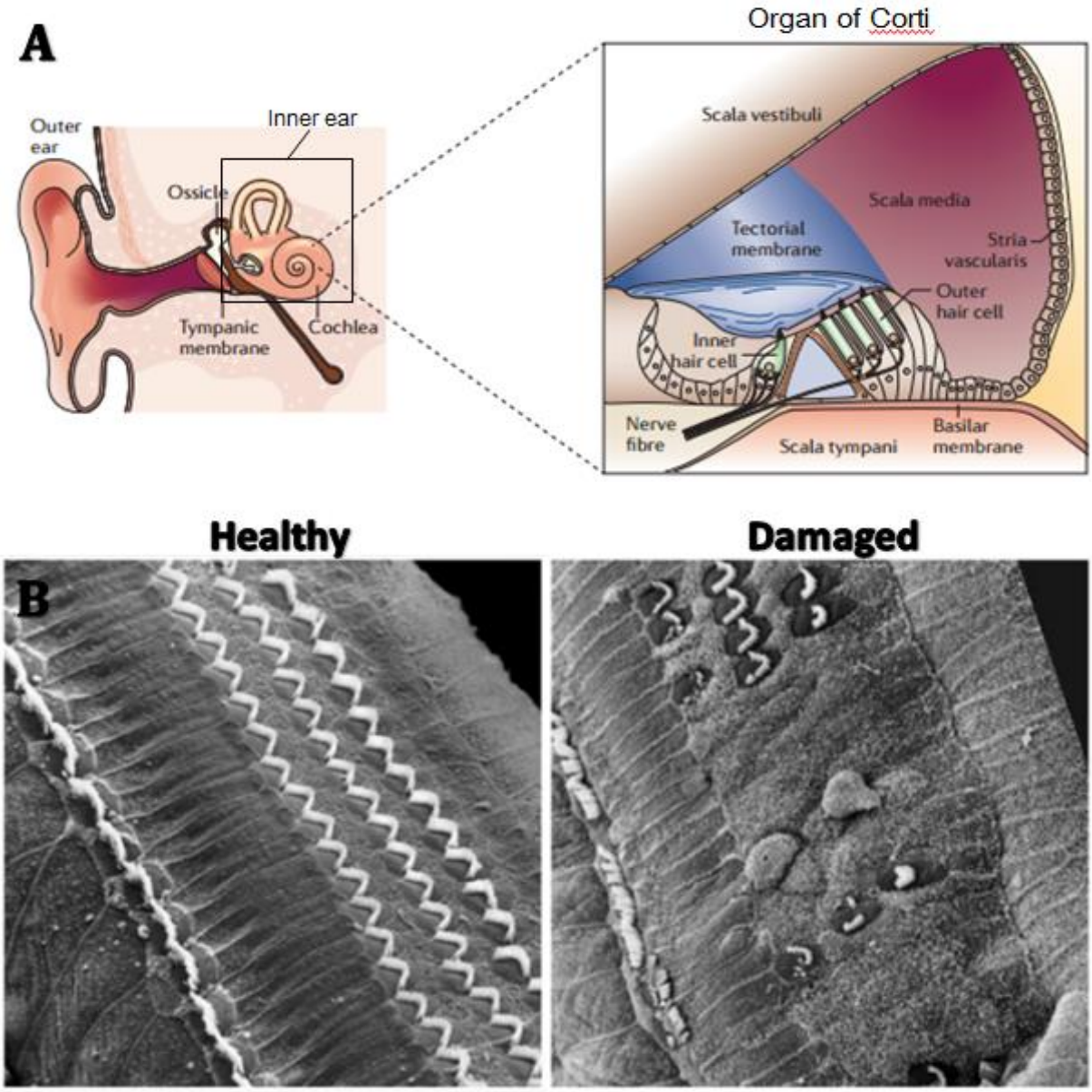


Figure 1: Physiology of Hearing

A, Schematic of human auditory sensory organ. Sound travels through as air pressure oscillations that induce vibrations in the inner ear. These vibrations are detected by hair cells in the organ of Corti of the Cochlea. HLD may occur from direct or indirect damage to auditory system including obstruction of vasculature supply, damaging of neural circuitry, or damaging of sensory hair cells. Diagram obtained and slightly modified from Muller, et al. 2015[44]. B, Electron micrograph images of intact cells (left) and damaged hair cells (right) from excessive noise exposure. Picture obtained from *House Ear Institute.*

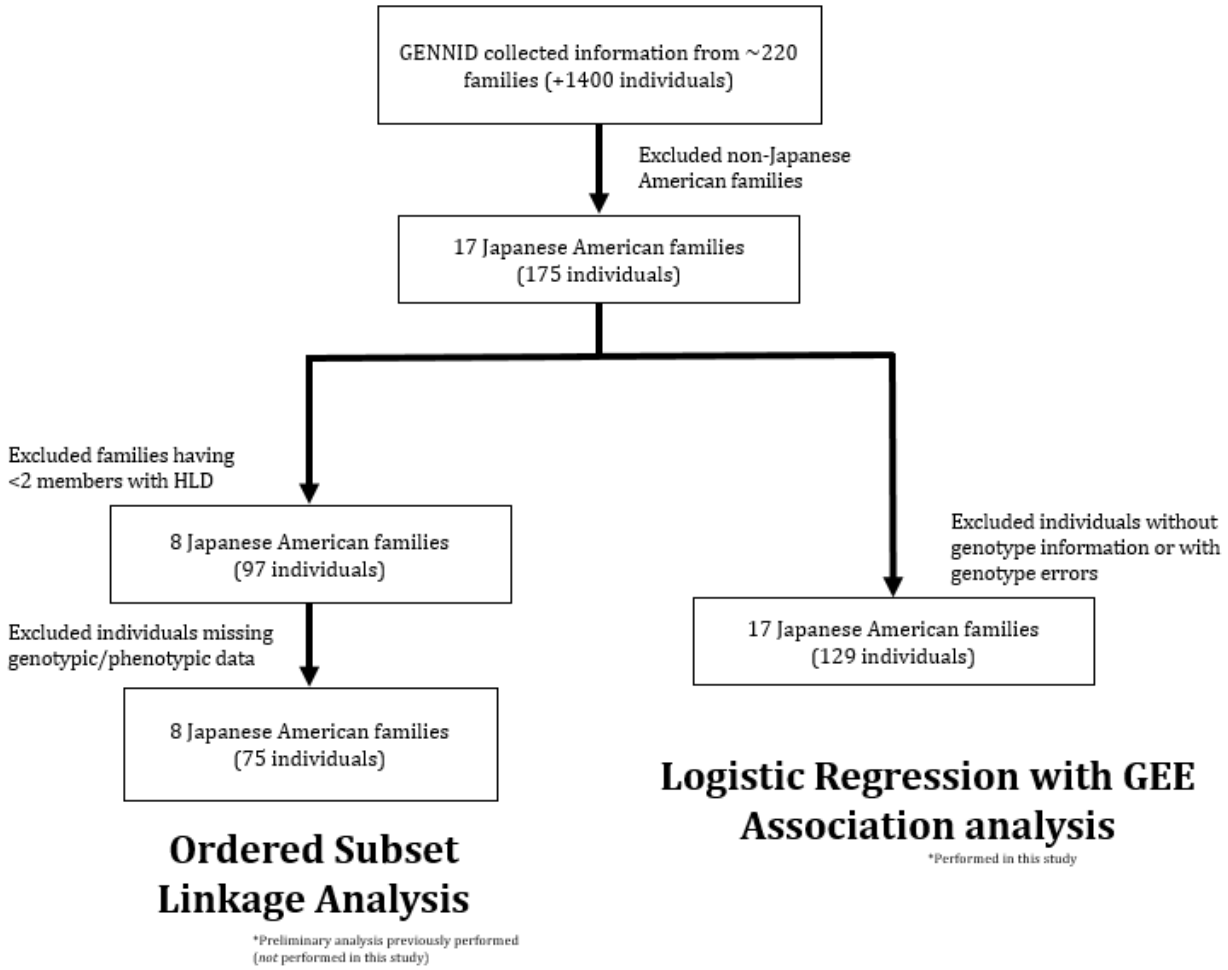


Figure 2: Flow Chart of participants who met inclusion/exclusion criteria

Subjects for this study were from the "Genetics of Non-insulin dependent diabetes mellitus" (GENNID) study, sponsored by the American Diabetes Association. OSA linkage analysis were not performed in this study, but in a previous analysis. Results from the linkage analysis help narrow the search space for the candidate gene association tests. The study focused in the Japanese American cohort of the GENNID study which contained the greatest number of HLD cases.

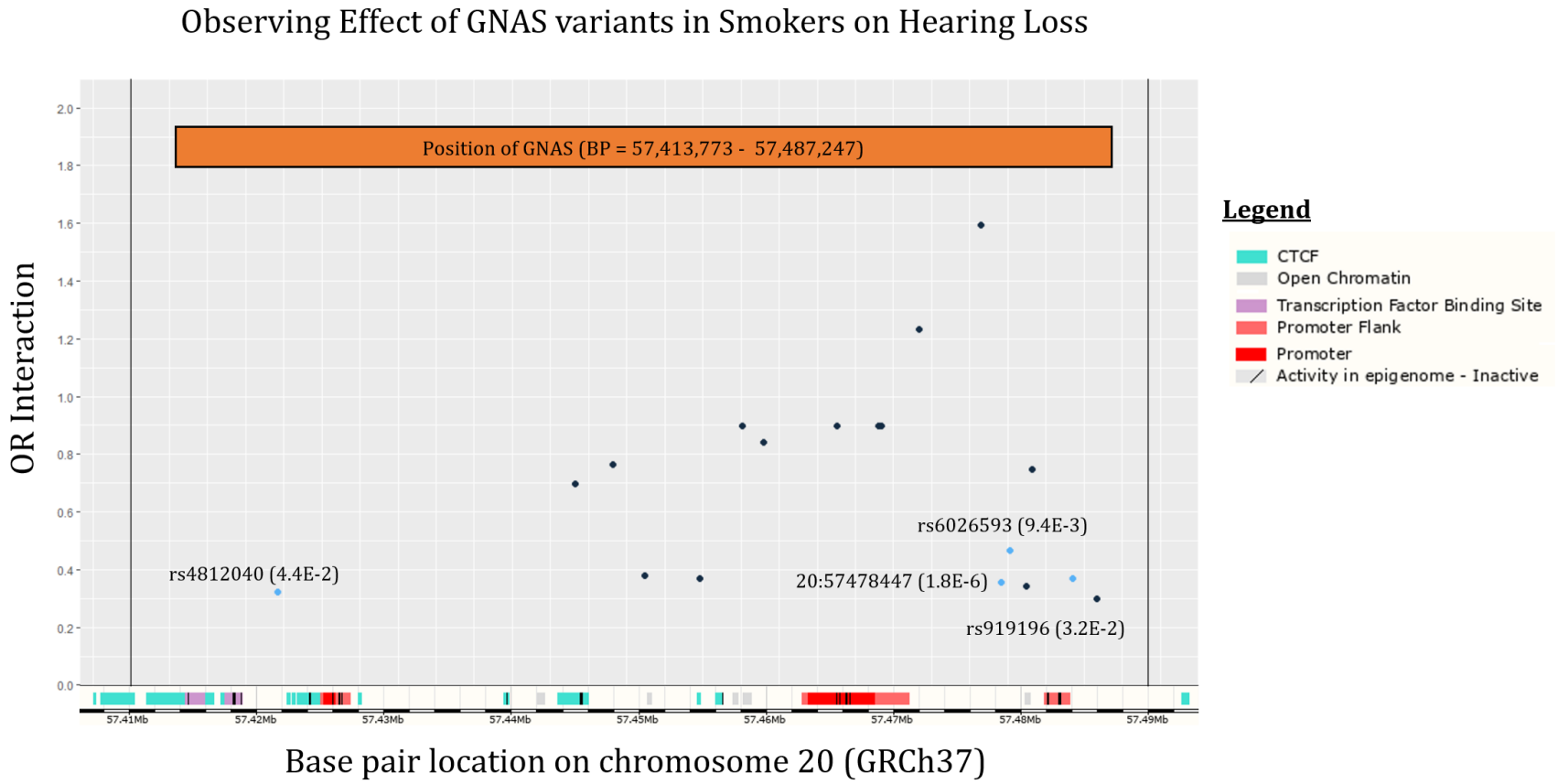


Figure 3: Observing effect of GNAS variants in Smokers on Hearing Loss

A total of four variants in GNAS (colored light blue) were significant with a P-value < 0.05. A majority of significant variants were situated at the upstream end of the gene, near GNAS's promoter and near the max LOD marker (0.3mb upstream). These variants are located in the intronic region suggesting its role in the regulation of GNAS expression. Gene structure information obtained from *Ensembl's regulatory build*.

References

1. WHO. Millions of People in the world have hearing loss that can be treated or prevented. *World Heal. Organ.* 20 (2013).
2. Lin, F. R., Niparko, J. K. & Ferrucci, L. Hearing Loss Prevalance in the United States. *Arch Otolaryngol Head Neck Surg.* **171,** 2011–2012 (2014).
3. Cruickshanks, K. J. *et al.* Prevalence of Hearing Loss in Older Adults in Beaver Dam , Wisconsin. *Am. J. Audiol.* **148,** 879–886 (1998).
4. Agrawal, Y., Platz, E. a & Niparko, J. K. Prevalence of Hearing Loss and Differences by Demographic Characteristics Among US Adults. *Arch. Intern. Med.* **168,** 1522–1530 (2008).
5. Lin, F. R., Thorpe, R., Gordon-salant, S. & Ferrucci, L. Hearing Loss Prevalence and Risk Factors Among Older Adults in the United States. 582–590 (2011). doi:10.1093/gerona/glr002
6. Cruickshanks, K. J. *et al.* Cigarette Smoking and Hearing Loss The Epidemiology of Hearing Loss Study. *Jama* **279,** 1715–1719 (1998).
7. Van Eyken, E., Van Camp, G. & Van Laer, L. The complexity of age-related hearing impairment: Contributing environmental and genetic factors. *Audiol. Neurotol.* **12,** 345–358 (2007).
8. Mills, J. H. & Going, J. A. Review of environmental factors affecting hearing. *Environ. Health Perspect.* **44,** 119–27 (1982).
9. Kvestad, E., Czajkowski, N., Krog, N. H., Engdahl, B. & Tambs, K. Heritability of hearing loss. *Epidemiology* **23,** 328–31 (2012).
10. Barone, J. A., Peters, J. M., Garabrant, D. H., Bernstein, L. & Krebsbach, R. Smoking as a

Risk Factor in Noise-induced Hearing Loss. *J. Occup. Environ. Med.* **29,** (1987).

11. Chang, J. *et al.* Effect of cigarette smoking and passive smoking on hearing impairment: Data from a population-based study. *PLoS One* **11,** 1–12 (2016).
12. Wild, D. C., Brewster, M. J. & Banerjee, A. R. Noise-induced hearing loss is exacerbated by long-term smoking. *Clin. Otolaryngol.* **30,** 517–520 (2005).
13. Yavuz, E. *et al.* Simultaneous measurement of electrocochleography and cochlear blood flow during cochlear hypoxia in rabbits. *J. Neurosci. Methods* **147,** 55–64 (2005).
14. Hauser, E., Watanabe, R. & Duren, W. Ordered subset analysis in genetic linkage mapping of complex traits. *Genetic* (2004).
15. Health, N. C. for C. D. P. and H. P. (US) O. on S. and. *The Health Consequences of Smoking—50 Years of Progress*. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General* (Centers for Disease Control and Prevention (US), 2014).
16. Fagerström, K. The epidemiology of smoking: health consequences and benefits of cessation. *Drugs* **62 Suppl 2,** 1–9 (2002).
17. Wetter, D. W. *et al.* Understanding the associations among education, employment characteristics, and smoking. *Addict. Behav.* **26,** 905–914 (2001).
18. Fagard, R. H. & Nilsson, P. M. Smoking and diabetes-The double health hazard! *Prim. Care Diabetes* **3,** 205–209 (2009).
19. Raffel, L. J. *et al.* The GENNID study: A resource for mapping the genes that cause NIDDM. *Diabetes Care* **19,** 864–872 (1996).
20. O'Connell, J. R. & Weeks, D. E. PedCheck: A program for identification of genotype

incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63,** 259–266 (1998).

21. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–11 (2001).

22. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. London B Biol. Sci.* **368,** (2013).

23. Smith, P. G. & Day, N. E. The design of case-control studies: The influence of confounding and interaction effects. *Int. J. Epidemiol.* **13,** 356–365 (1984).

24. Yu, W., Clyne, M., Khoury, M. J. & Gwinn, M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* **26,** 145–6 (2010).

25. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33,** D514–D517 (2004).

26. Ballinger, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. doi:10.1177/1094428104263672

27. Suktitipat, B. *et al.* The robustness of generalized estimating equations for association tests in extended family data. *Hum. Hered.* **74,** 17–26 (2012).

28. Qayyum, R. *et al.* Genome-wide association study of platelet aggregation in African Americans. *BMC Genet.* **16,** 58 (2015).

29. Chiu, Y.-F. *et al.* Identification of rare variants for hypertension with incorporation of linkage information. *BMC Proc.* **8,** S109 (2014).

30. Cupples, L. A. *et al.* The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med. Genet.* **8**

Suppl 1, S1 (2007).

31. Underbjerg, L., Sikjaer, T., Mosekilde, L. & Rejnmark, L. Pseudohypoparathyroidism - epidemiology, mortality and risk of complications. *Clin. Endocrinol. (Oxf).* **84,** 904–911 (2016).

32. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47,** 435–444 (2015).

33. Blackman-Mack, R., Peters, C., Langham, S., Rajput, K. & Sirimanna, T. An assessment of auditory function in infants with congenital hypothyroidism. *Endocr. Abstr.* (2015). doi:10.1530/endoabs.39.OC6.6

34. Ng, L. *et al.* Hearing loss and retarded cochlear development in mice lacking type 2 iodothyronine deiodinase. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3474–9 (2004).

35. Dror, A. A. *et al.* Atrophic thyroid follicles and inner ear defects reminiscent of cochlear hypothyroidism in Slc26a4-related deafness. *Mamm. Genome* **25,** 304–316 (2014).

36. Belin, R. M., Astor, B. C., Powe, N. R. & Ladenson, P. W. Smoke Exposure Is Associated with a Lower Prevalence of Serum Thyroid Autoantibodies and Thyrotropin Concentration Elevation and a Higher Prevalence of Mild Thyrotropin Concentration Suppression in the Third National Health and Nutrition Examination Survey (NHANES III). *J. Clin. Endocrinol. Metab.* **89,** 6077–6086 (2004).

37. Wiersinga, W. M. Smoking and thyroid. *Clin. Endocrinol. (Oxf).* **79,** 145–151 (2013).

38. Hickson, V. M. Improvement in neonatal hearing loss following treatment with thyroxin. *Arch. Dis. Child.* **96,** A31–A32 (2011).

39. Khan, S. Y. *et al.* Mutations of the *RDX* gene cause nonsyndromic hearing loss at the

DFNB24 locus. *Hum. Mutat.* **28,** 417–423 (2007).

40. Kitajiri, S. *et al.* Radixin deficiency causes deafness associated with progressive degeneration of cochlear stereocilia. *J. Cell Biol.* **166,** 559–70 (2004).

41. Khan, S. Y. *et al.* Mutations of the RDX gene cause nonsyndromic hearing loss at the DFNB24 locus. *Hum. Mutat.* **28,** 417–23 (2007).

42. Smith, R. J., Shearer, A. E., Hildebrand, M. S. & Van Camp, G. *Deafness and Hereditary Hearing Loss Overview*. *GeneReviews(®)* (University of Washington, Seattle, 1993).

43. Clark, J. Uses and abuses of hearing loss. *ASHA* **23,** 493–500 (1981).

44. Müller, U. & Barr-Gillespie, P. G. New treatment options for hearing loss. *Nat. Rev. | Drug Discov.* **14,** 346–365 (2015).

APPENDIX

APPENDIX A: Linkage evidence of gene-by-environment interaction in subset of families

* The fraction indicates the numerator as the top number of families from high to low ranks based on percentage of smokers per-family ranking (see Table 2) over the denominator which is the total number of 8 families. This subset of families was found to yield significant evidence gene-by-smoking interaction (P<0.05).

Smoking covariate	**CHR**	**Marker at OSA max LOD**	**Base Pair location** (GRCH37/hg 19)	**OSA Max LOD**	**Original Max LOD**	**P-value**	**#Families / Total***
Packs-per-day	**8q21.3**	D8S1119	87172020	3.54	1.83	0.04	4/8
Ever smoking	**7p21.1**	D7S3051	18,284,528	3.3	1.77	0.04	2/8
	11q23.1	D11S1986	111,223,605	2.1	1.12	0.03	2/8
	12q32.1	D12S1045	130,397,738	2.6	1.15	0.01	3/8
	15q26.2	D15S816	95,019,784	2.9	1.42	0.01	5/8
	20q13.3	D20S171	57,808,030	3	1.77	0.02	3/8

APPENDIX B: Interpreting Logistic Models with Interactions

The multivariable logistic regression is expressed as followed:

$$\text{Logit}\left(\text{Probability of HLD}_{ij}\right) = \beta_0 + \beta_1\left(\text{Smoking}_{ij}\right) + \beta_2\left(\text{Genotype}_{ij}\right) + \beta_3\left(\text{Smoking}_{ij} * \text{Genotype}_{ij}\right) + \beta_4(\text{Age}_{ij})$$

The interaction term allows the odds ratio of smoking to depend on genotype status. Since we are interested in evaluating whether the relation of smoking and HLD is dependent on genotype, we focus on these variables (Z represents age, sex, diabetes, and education status). Genotype (G) is a binary trait (aa = 1, AA or Aa = 0) and smoking (S) is a continuous variable. Below describes how to interpret parameters from logistic regression to interpret the main effect of smoking and genotype and their interaction in odds ratio.

Smoking main effect – Interpreting β_1 coefficient (G = 0):

For smoking = *k* (any given value of packs/day except 0)

$$\text{Log odds of HLD} = \beta_0 + \beta_1(k) + \beta_2(0) + \beta_3(k * 0) + \text{Z}$$
$$\text{odds of HLD} = e^{\beta_0+\beta_1(k)+\text{Z}}$$

For smoking = *k* + 1

$$\text{Log odds of HLD} = \beta_0 + \beta_1(\text{k}+1) + \beta_2(0) + \beta_3(k * 0) + \text{Z}$$
$$\text{odds of HLD} = e^{\beta_0+\beta_1(k+1)+\text{Z}}$$
$$\text{odds of HLD} = e^{\beta_0+\beta_1(k)+\beta_1+\text{Z}}$$

Therefore, the odds ratio (going from *smoking = k* to *smoking = k+1*) is:

$$OR = \frac{\text{odds when smoking} = \text{k}+1}{\text{odds when smoking} = \text{k}} = \frac{e^{\beta_0+\beta_1(k)+\beta_1+\text{Z}}}{e^{\beta_0+\beta_1(k)+\text{Z}}} = e^{\beta_1}$$

For every unit increase in packs/day, the odds of HLD increases by a factor of $e^{\beta_1+\text{Z}}$ in individuals that do not carry the risk genotype. This is the main effect of smoking.

Genotype main effect - Interpreting β_2 coefficient (when S = 0):

For genotype = 0 (AA or Aa)

$$\text{Log odds of HLD} = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0 * 0) + \text{Z}$$
$$\text{odds of HLD} = e^{\beta_0+\text{Z}}$$

For genotype = 1 (aa)

$$\text{Log odds of HLD} = \beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0 * 1) + Z$$
$$\text{odds of HLD} = e^{\beta_0+\beta_2+Z}$$

Therefore, the odds ratio going from *genotype = 0* to *genotype = 1* in nonsmokers is:

$$OR = \frac{\text{odds when genotype(aa)} = 1}{\text{odds when smoking(AA or Aa)} = 0} = \frac{e^{\beta_0+\beta_2+Z}}{e^{\beta_0+Z}} = e^{\beta_2}$$

The odds of HLD is e^{β_2+Z} in individuals carrying the risk genotype relative to individuals not carrying the risk genotype. This is the main effect of genotype

Smokers with risk genotype – Interpreting interaction (when G = 1, S ≠ 0)

For smoking = k (any given value of packs/day)

$$\text{Log odds of HLD} = \beta_0 + \beta_1(k) + \beta_2(1) + \beta_3(1 * 1) + Z$$
$$\text{odds of HLD} = e^{\beta_0+\beta_1(k)+\beta_2+\beta_3+Z}$$

Thus the odds ratio going from *genotype = 0* to *genotype = 1* in smokers is:

$$OR = \frac{\text{odds when smoking } (k) \text{and genotype } (+)}{\text{odds when smoking } (k) \text{ and genotype } (-)} = \frac{e^{\beta_0+\beta_1(k)+\beta_2+\beta_3+Z}}{e^{\beta_0+\beta_1(k)+Z}} = e^{\beta_2+\beta_3}$$

In smokers, the odds of HLD is $e^{\beta_2+\beta_3+Z}$ in individuals carrying the risk genotype relative to individuals not carrying the risk genotype. In the model, $e^{\beta 3}$ relates the $OR_{\text{smokers + genotype (-)}}$ to the $OR_{\text{smokers + genotype (+)}}$. If $\beta_3 = 0$, then the two ORs are identical and there is no interaction. This is the genotype-by-smoking effect.