



HAL
open science

A genome-wide association study for age-related hearing impairment in the Saami

Guy van Camp, Lut van Laer, Jeroen Huyghe, Samuli Hannula, Els van Eyken, Dietrich Stephan, Elina Mäki-Torkko, Pekka Aikio, Alana Lysholm-Bernacchi, Martti Sorri, et al.

► **To cite this version:**

Guy van Camp, Lut van Laer, Jeroen Huyghe, Samuli Hannula, Els van Eyken, et al.. A genome-wide association study for age-related hearing impairment in the Saami. *European Journal of Human Genetics*, 2010, n/a (n/a), pp.n/a-n/a. 10.1038/ejhg.2009.234 . hal-00501993

HAL Id: hal-00501993

<https://hal.science/hal-00501993>

Submitted on 13 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 A genome-wide association study for age-related hearing impairment in the Saami

2

3 Lut Van Laer^{1,‡}, Jeroen R Huyghe^{1,‡}, Samuli Hannula², Els Van Eyken¹, Dietrich A Stephan³,
4 Elina Mäki-Torkko², Pekka Aikio⁴, Erik Fransen¹, Alana Lysholm-Bernacchi^{3,†}, Martti Sorri²,
5 Matthew J Huentelman³, Guy Van Camp^{1,*}

6

7 [‡]The first two authors contributed equally to the work

8 [†]In loving memory

9

10 ¹Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; ²Department of
11 Otorhinolaryngology, University of Oulu, Oulu, Finland; ³Neurogenomics Division, The
12 Translational Genomics Research Institute, Phoenix, Arizona, USA; ⁴Thule
13 Institute, University of Oulu, Oulu, Finland

14

15 *Correspondence:

16 Prof. Dr. Guy Van Camp

17 Department of Medical Genetics

18 University of Antwerp, Campus Drie Eiken

19 Universiteitsplein 1, B-2610 Antwerp, Belgium

20 Tel.: +323/820.24.91

21 Fax: +323/820.25.66

22 Email: guy.vancamp@ua.ac.be

23

24 **Running title** GWAS for ARHI in the Saami

25 This study aimed to contribute to the elucidation of the genetic basis of age-related
26 hearing impairment (ARHI), a common multifactorial disease with an important genetic
27 contribution as demonstrated by heritability studies. We conducted a genome-wide
28 association study (GWAS) in the Finnish Saami, a small ancient genetically isolated
29 population without evidence of demographic expansion. The choice of this study
30 population was motivated by its anticipated higher extent of linkage disequilibrium,
31 potentially offering a substantial power advantage for association mapping. DNA
32 samples and audiometric measurements were collected from 352 Finnish Saami
33 individuals, aged between 50 and 75 years. To reduce the multiple testing burden, we
34 applied principal component analysis to the multivariate audiometric phenotype. The
35 first three principal components (PC) captured 80% of the variation in hearing
36 thresholds, while maintaining biologically important audiometric features. All subjects
37 were genotyped with the Affymetrix 100K chip. To account for multiple levels of
38 relatedness among subjects, as well as population stratification, association testing was
39 performed using a mixed model. We summarized top-ranking association signals for the
40 three traits under study. The top-ranked SNP, rs457717 (P-value 3.55×10^{-7}), was
41 associated with PC3 and was localised in an intron of the IQ motif-containing GTPase-
42 activating-like protein (*IQGAP2*). Intriguingly, SNP rs161927 (P-value 0.000149),
43 seventh-ranked for PC1, was positioned immediately downstream of the metabotropic
44 glutamate receptor 7 gene (*GRM7*). As a previous GWAS in a European and a Finnish
45 sample set already suggested a role for *GRM7* in ARHI, this study provides further
46 evidence for the involvement of this gene.

47

48 **Key words:** Saami – isolated population – mixed model - genome-wide association study -
49 age-related hearing impairment - presbycusis

50 **Introduction**

51 Age-related hearing impairment (ARHI), or presbycusis, is the most frequent sensory
52 disability among the elderly.^{1,2} The involvement of genetic factors in the development of this
53 multifactorial disorder was demonstrated by several heritability studies.³⁻⁵ Only recently the
54 first studies aiming at the identification of these genetic factors have been carried out.
55 Hitherto, three linkage studies have been performed,⁶⁻⁸ but only one region on chromosome
56 8q24.13-q24.22 reached genome-wide significance.⁸ In addition, several association studies
57 based on a candidate gene approach and two genome-wide association studies (GWAS)^{8,9}
58 have been conducted. Early candidate gene-based association studies are reviewed in Van
59 Eyken et al.¹⁰ Recently, based on a candidate gene approach, we reported associations
60 between ARHI and variants in *GRHL2*,¹¹ encoding a transcription factor that is also involved
61 in monogenic hearing loss. In addition, based on a pooling GWAS, we reported association
62 with variants in *GRM7*,⁹ encoding a metabotropic glutamate receptor assumed to modulate
63 hair cell excitability and synaptic efficacy. Acquired mitochondrial mutations have also been
64 proposed as a cause of ARHI in humans. These include both the “common” 4977 bp
65 mitochondrial DNA deletion associated with ageing¹² and an accumulation of many different
66 mitochondrial mutations.¹³

67

68 Despite these first series of studies, the genetic etiology of ARHI remains elusive. In this
69 paper we present the results of a GWAS for ARHI in the Finnish Saami. The Saami,
70 previously known as Lapps, live across Northern Scandinavia and the Kola Peninsula.
71 Traditionally, Saami have been fishers, hunters and reindeer herders. A minority of the Saami
72 continues to be engaged in these traditional economic activities, while the majority has
73 adopted other occupations and a more westernised lifestyle. Among European populations,
74 the Saami are considered a population genetic “outlier”. Despite the fact that both the Saami

75 languages and Finnish belong to the Uralic language group, there is strong genetic
76 differentiation between Saami and Finns.¹⁴ Several genetic studies have been conducted to
77 shed light on the genetic origin of the Saami.¹⁵⁻¹⁸ These studies suggested that the Saami are
78 descendants of Europeans that, several thousand years ago, reached northern Fennoscandia
79 from both a western and an eastern route and subsequently got admixed. They also found a
80 small eastern Asian contribution to the gene pool.^{18,19} Several genetic studies further have
81 suggested that the extent of linkage disequilibrium (LD) in the Saami is dramatically higher
82 compared to outbred populations.²⁰⁻²³ Higher levels of LD are observed in other population
83 isolates too.^{24,25} Most of the genetic studies on genetically isolated populations were
84 conducted on young founder populations, in which the high extent of LD can be attributed to
85 a founder effect. In ancient small genetically isolated populations without evidence of
86 demographic expansion, however, high levels of LD are believed to have been generated by
87 genetic drift.^{26,27} The Saami represent such a population. Therefore, the choice of this study
88 population was based on its anticipated statistical power advantage for association mapping,
89 offering a better genomic coverage for a given set of markers, compared to outbred
90 populations.

91 Due to the expected higher levels of LD in the Saami population and based on what was
92 known about the structure of the humane genome at the time this study was designed, we
93 anticipated that the Affymetrix 100K array pair would entail sufficient genomic coverage for
94 the identification of putative ARHI susceptibility genes.

95

96

97 **Materials and methods**

98 **Subjects** Saami subjects, aged between 50 and 75 years, were enrolled through the public
99 population register from four municipalities in northern Finland (Ivalo, Utsjoki, Vuotso and

100 Enontekiö) in three stages. In a first stage, geographical criteria were applied; only subjects
101 from an area with a high probability of including Saami were selected. In a second stage,
102 putative study subjects were invited based on an evaluation made by an expert in Saami
103 Communities. Finally, the Saami identity of the subject in question was confirmed in a
104 discussion taking place between the subject and the researcher. Except for the age range that
105 was used, ARHI selection criteria were as described previously.²⁸ Briefly, all volunteers
106 completed an extended questionnaire detailing medical history and exposure to environmental
107 factors putatively influencing hearing. Subjects suffering from one or more medical
108 pathologies from a standardized list of conditions that could potentially affect hearing ability
109 were excluded. All subjects who passed this initial screening underwent an otoscopic
110 investigation before pure tone audiometry. Air conduction thresholds were measured at 0.125,
111 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz, and bone conduction at 0.5, 1, 2 and 4 kHz for all
112 participating volunteers. Audiological exclusion criteria were an air-bone gap of more than 15
113 dB averaged over 0.5, 1 and 2 kHz in at least one ear or asymmetrical hearing impairment
114 with a difference in air conduction thresholds exceeding 20 dB in at least 2 frequencies
115 between 0.5, 1 and 2 kHz. Written informed consent was obtained from all volunteers. The
116 study was approved by the Finnish National Advisory Board on Health Care Ethics and by the
117 ethical committees or the appropriate local institutional review boards at all participating
118 institutions.

119

120 **Description of the phenotype** For each subject, only hearing thresholds from the best ear,
121 defined as the ear with the lowest average threshold over all measured frequencies, were used
122 in the analyses. Supplementary Figure 1 shows that hearing thresholds increase with age and
123 that, for the higher frequencies, thresholds for males are higher and display greater variability
124 (i.e. there is heteroscedasticity) than for females. We therefore adjusted hearing thresholds for

125 age and sex. The effect of age was removed by regressing the log-transformed threshold
126 values ($\log_{10}(20+\text{threshold})$) for each of the measured frequencies on age, age squared and
127 age cubed. To resolve the heteroscedasticity problem and to correct for sex, this adjustment
128 for age was carried out in males and females separately, after which the residuals of each
129 regression were scaled and subsequently combined. Note that adding the age and sex
130 covariates in the model for association testing would not deal with this problem satisfactorily.
131 To deal with the multivariate audiometric phenotype, we applied principal component
132 analysis as a data reduction technique. Performing association tests for each frequency
133 separately would have resulted in a large number of tests, worsening the multiple testing
134 problem. In addition, such analysis would completely have ignored the correlation structure of
135 the audiometric data, i.e. the shape of the audiogram. The latter would also apply when simple
136 pure-tone averages (PTAs), which are more widely used in hearing research, would have been
137 analysed. Our first principal component (PC1) was highly correlated with the PTA over all
138 frequencies, adjusted for sex, age, age square and age cubed (Spearman correlation = 0.96).
139 Spearman correlations between PC1 and sex and age adjusted PTAs for the high frequencies
140 (2 kHz to 8 kHz) and low frequencies (0.125 kHz to 1 kHz) were 0.88 and 0.61, respectively.
141 The association testing results for PC1, therefore, would not differ much from those that
142 would be obtained when PTAs would have been analysed.
143 Table 1 lists the eigenvalues and eigenvectors for the first three principal components (PCs),
144 PC1 to PC3. Results of calculations on males and females separately yielded similar values,
145 justifying data-pooling (results not shown). Each of the PCs provides independent bits of
146 information. The first three PCs were retained for the analysis as they captured approximately
147 80% of the total variation. The coefficients of the eigenvectors (loadings) are directly
148 proportional to the correlations between the transformed hearing thresholds corrected for age
149 and gender, and each of the PCs (see e.g. Johnson and Wichern²⁹). PC1 is clearly a ‘size’

150 variable, providing an overall measure of a subject's hearing ability. PC2 and PC3 are 'shape'
151 variables. PC2 contrasts low with high frequencies and is a measure for the slope of the
152 audiogram. PC3 contrasts the middle frequencies with the lower and higher frequencies and
153 can be considered a measure for the concavity of an audiogram (Table 1 and Figure 1). These
154 phenotypes are biologically meaningful^{30,31} and we have previously shown that they have
155 significant heritabilities.⁸

156

157 **Genotyping and Quality Control of the data** Genomic DNA was extracted using standard
158 protocols and diluted to 50 ng/ μ l. The quality of the genomic DNA was assessed by
159 spectrophotometric analysis and gel electrophoresis. Each sample was genotyped on the
160 GeneChip[®] Human Mapping 100K array pair from Affymetrix (Santa Clara, CA, USA) as
161 described in the GeneChip[®] Human Mapping 100K Assay Manual (Affymetrix, Santa Clara,
162 CA, USA). Genotype calling was performed by the BRLMM algorithm of the Affymetrix
163 GeneChip Genotyping Analysis Software (GTYPE) version 4.1. Quality control of the data
164 was performed based on the following criteria: a sample call rate of > 94%, a marker call rate
165 of > 90%, a Hardy-Weinberg equilibrium P-value of > 0.001 and a minor allele frequency of
166 > 5%. In total 352 Saami samples were genotyped with the Affymetrix 100K array. One
167 subject was excluded because of a low sample call rate. The total genotyping rate in the
168 remaining samples was 0.9931. One additional sample was removed due to an unintentional
169 duplication event, and three further subjects were excluded due to missing phenotypic
170 information. After data quality control, 347 subjects and 83,381 SNPs were taken forward for
171 statistical analysis. Genotyping for SNP rs11928865 was performed using a Taqman assay
172 (Applied Biosystems, Foster City, CA, USA) on a Lightcycler480 realtime PCR machine
173 (Roche, Basel, Switzerland).

174

175 **Association testing** Estimates of genome-wide pairwise identity-by-descent sharing, provided
176 by PLINK,³² revealed a substantial degree of undocumented relatedness among subjects. This
177 was also reflected in an excess of SNPs that were not in Hardy-Weinberg equilibrium. The
178 results of a linear regression for each of the three quantitative traits would, therefore, not be
179 trustworthy as dependent data lead to underestimation of the standard errors for the regression
180 coefficients, and hence, *P*-values that are too small. In addition, population stratification may
181 confound the results of the analysis. To account for the dependency in the data as well as
182 population stratification, we analysed the data using a mixed model.³³ With this approach the
183 covariance matrix of the phenotype is modelled as:

$$184 \quad \text{Var}(y) = 2K\sigma_g^2 + I\sigma_e^2$$

185 where *y* is the column vector with the phenotypes of the 347 subjects, *K* is a matrix with
186 kinship coefficients estimated based on the genome-wide data, *I* is an identity matrix and σ_g^2
187 and σ_e^2 are the genetic and the residual variance respectively. This was implemented using the
188 MIXED procedure of SAS 9.1.3 within a SAS macro, which was based on that used by Yu *et*
189 *al.*³³ The *2K* matrix with the proportion of alleles shared identical-by-descent, was estimated
190 using PLINK.³² The model also corrected for origin. As samples originated from 4
191 municipalities, 3 dummy variables were included as covariates in the model. An additive
192 genetic model was assumed for each SNP. As it was computationally infeasible to run this
193 analysis for the entire set of SNPs, for each trait only the top 4000 SNPs (approximately 5%
194 of the total number of SNPs passing quality control) resulting from an initial screening based
195 on a linear regression analysis (naive analysis) went to the mixed model analysis. To verify
196 that no true association signals would be lost by using this analytical strategy, we first
197 examined how well the *P*-values of the naive analysis correlated with those of the mixed
198 model analysis. For this purpose, we analysed 8000 randomly selected SNPs meeting quality
199 control standards, representing approximately 10% of the SNPs that passed quality control,

200 with the mixed model. Comparison of P -values between naive analysis and mixed model
201 results showed that no SNPs with highly significant P -values in the mixed model analysis
202 showed a non-significant P -value in the naive analysis. A reciprocal relationship would be
203 attributed to inflated test statistics. Therefore, for the data at hand, by analysing the top 5%
204 (4000 SNPs) of a naive linear regression analysis, the chance of missing any SNPs that would
205 not be significant in the naive analysis but highly significant in the mixed model analysis
206 seems very low (Supplementary Figure 2). Quantile-quantile plots in Supplementary Figure 2
207 show that residual population stratification or cryptic relatedness are of minor concern.
208 Genomic control inflation factors varied from 1.02 to 1.08. The graph at the right of each
209 panel in Supplementary Figure 2 further shows that P -values obtained from the naive analysis
210 tend to be too low. For example, for PC1, the most extreme $-\log_{10}(P\text{-value})$ from the naive
211 analysis is 4, while this is 3.5 for the mixed model.

212

213

214 **Results**

215 After quality control, 347 subjects and 83,381 SNPs were taken forward for association
216 testing. The quantitative traits tested were the first three principal components (PCs)
217 calculated from the log-transformed hearing threshold values, adjusted for age and sex. To
218 account for the multiple levels of relatedness present in our data and potential population
219 stratification, associations between SNPs and the traits were tested using a mixed model
220 approach, based upon genome-wide kinship estimates. Due to computational constraints, only
221 the top 4000 SNPs (approximately 5% of the total number of SNPs passing quality control)
222 resulting from an initial screening based on a linear regression analysis went to the mixed
223 model analysis. In the methods section we have outlined that, using this strategy, it is unlikely

224 that strong association signals will be lost. For each of the three PCs, the top 25 SNPs
225 resulting from this analysis are listed in Tables 2 to 4.

226 If, in agreement with other GWAS,³⁴ we used a P -value cutoff of 5×10^{-7} to declare genome-
227 wide significance, the top-ranked SNP for PC3 (rs457717; Table 4) survived this significance
228 threshold (P -value = 3.55×10^{-7}). This SNP is located in an intron of the IQ motif-containing
229 GTPase-activating-like protein (*IQGAP2*). The second ranked SNP (rs1697845) is also
230 positioned in *IQGAP2*, and is highly correlated with the first ($D' = 1$; $r^2 = 0.965$), indicating
231 that this signal is not a genotyping artefact.

232 For the other two traits, PC1 and PC2, none of the SNPs reached this level of significance
233 (Table 2 and 3). Intriguingly, however, the eight-ranked SNP for PC1 (rs161927; Table 2; P -
234 value 0.000149) was positioned 40 kb downstream of the metabotropic glutamate receptor 7
235 gene (*GRM7*; Supplementary Figure 3). We have previously reported an association of a
236 common variant in the *GRM7* gene, rs11928865, with hearing ability based on a pooling-
237 based GWAS in a European sample set.⁹ Therefore, we also genotyped this SNP in the Saami
238 and analysed this data using the mixed model approach. This resulted in a P -value of 0.045.
239 However, the A-allele was associated with an elevated hearing threshold in the Saami sample
240 set, while in the European sample set the associated allele was T.⁹ One of the two top-scoring
241 *GRM7* SNPs resulting from a GWAS in a Finnish sample set,⁹ rs779701, was present on the
242 100K Affymetrix array. In the Saami sample set this SNP did, however, not result in a
243 significant association (P -value of 0.735). The LD-pattern within the *GRM7* region deduced
244 for the Saami (Supplementary Figure 3D) was similar to that observed in the HapMap for the
245 European population (Supplementary Figure 3C). No obvious biological meaning related to
246 the inner ear could be deduced for the gene regions of the seven higher-ranked SNPs for PC1.
247 For PC2 no strong biological candidates were present among the top 25 ranking SNPs (Table
248 3). The presence of six SNPs from the same gene (KH domain-containing RNA-binding

249 signal transduction-associated protein 2; *KHDRBS2*) within Table 3 is due to the fact that all
250 of these SNPs reside in the same region of high LD.

251

252 **Discussion**

253 We conducted a genome-wide association scan for ARHI traits derived from audiometric data
254 in the genetically isolated Finnish Saami population. We summarised the information
255 contained in each audiogram by the first three components obtained with a principal
256 component analysis. Previously we have shown that these traits have significant
257 heritabilities.⁸ By using these audiometric measures we retain biologically important
258 information on the shape of the audiogram that would be lost when simple pure-tone averages
259 or individual frequencies would be analysed. Indeed, premortem-derived audiometric patterns
260 have been correlated to pathophysiological findings in human temporal bones.^{30,31} Nelson et
261 al.³⁵ studied individuals with downward sloping audiograms and reported that the severity of
262 hearing loss, based on audiometric thresholds, was highly associated with a decrease in stria
263 vascularis volume and the degeneration of outer and inner hair cell and ganglion cell
264 populations. The slope of the audiogram was significantly associated with the extent of
265 ganglion cell degeneration.

266

267 Our association scan resulted in two SNPs in or near genes that are reasonable biological
268 candidates and that are candidates for further research. The top-ranked SNP, rs457717 (*P*-
269 value 3.55×10^{-7}), that reached genome-wide significance, was associated with PC3 and was
270 localised in an intron of the IQ motif-containing GTPase-activating-like protein (*IQGAP2*).
271 *IQGAP2* is a member of the Ras superfamily of GTPases. This class of proteins regulates a
272 wide variety of cellular signalling pathways. *IQGAP2* is expressed inside the cochlea³⁶ and
273 has been implicated in cadherin-mediated cell adhesion.³⁷ As cadherin 23 (*CDH23*) plays an
274 important role in the sensory hair cells³⁸ and mutations in *CDH23* are responsible for
275 syndromic and nonsyndromic autosomal recessive deafness in humans^{39,40} and age-related
276 hearing loss in inbred mouse strains,⁴¹ this warrants further research. Intriguingly, SNP

277 rs161927 (*P*-value 0.000149) that ranked eight for PC1, was positioned immediately
278 downstream of the metabotropic glutamate receptor 7 gene (*GRM7*). A previous GWAS in a
279 European and a Finnish sample set that was based on an age and gender-corrected *Z*-score
280 highly correlated with our PC1,^{9,11,28,42} already suggested a role for *GRM7* in ARHI.⁹ *GRM7*
281 encodes a metabotropic glutamate receptor which can be activated by L-glutamate, the
282 primary excitatory neurotransmitter in the cochlear hair cell. *GRM7* expression was observed
283 in inner and outer hair cells and in the spiral ganglion nerve cell bodies of the inner ear.⁹ We
284 have hypothesized that certain *GRM7* variants may cause an increased susceptibility to
285 glutamate excitotoxicity, which may consequently lead to the development of ARHI.
286 However, variants that were significant in our previous study did not replicate in our current
287 study. The LD structure within the *GRM7* region indicates that these are independent signals.
288 These results, therefore, are consistent with allelic heterogeneity which is a plausible scenario
289 given the likely absence of any selective pressure for this disorder. Independent studies are
290 needed to investigate the role of common variants in *GRM7* in ARHI.

291

292 This study was carried out in the genetically isolated Saami population. A number of studies
293 provide evidence that the extent of LD is much higher in the Saami compared to other
294 European populations.²⁰⁻²³ This would provide an important power advantage, as higher LD
295 leads to better genomic coverage. It has been proposed that association studies based on a
296 map of modest marker density within a population isolate such as the Saami, would provide a
297 first rough identification of the region associated with a particular complex disease. This first
298 demarcation of the region should then preferably be refined by association studies in an
299 outbred population using a denser map.²² Elsewhere, we will publish the results of an
300 elaborate genome-wide evaluation of the relative power for GWAS in the Saami.

301 A possible limitation of this study is the limited sample size. Recent studies indicate that
302 several thousand subjects need to be recruited in order to obtain sufficient power to detect the
303 effect sizes typically expected for complex phenotypes (with each contributing variant
304 explaining as little as 0.1 to 1.3% of the variation),^{43,44} although successful GWAS have been
305 conducted using only 96 cases and 50 controls.⁴⁵ Obviously, large sample sizes can never be
306 recruited in small isolated populations such as the Saami. A specific difficulty of working
307 with isolated populations became apparent in the current study. Due to the fact that isolated
308 populations are often derived from a limited number of founders and that inbreeding is
309 sometimes inevitable, an increased relatedness is observed within such populations. As a
310 consequence, classical statistical methods that assume unrelatedness, cannot be applied to
311 analyse an association study performed in a population isolate. In our exploratory phase we
312 investigated several approaches including a naive linear regression analysis and a principal
313 components-based correction (EIGENSTRAT⁴⁶). However, for the latter approach the basic
314 assumptions were violated and the inbreeding and relatedness present in our Saami sample
315 created spurious axes of variation.⁴⁷ Mixed models offered a solution. In this approach we
316 estimated genetic relationships between study subjects based on genome-wide data and
317 subsequently accounted for them in the analysis.³³

318

319 In conclusion, we present the results of a first GWAS performed in the genetically isolated
320 Saami population. We identified *IQGAP2* as a candidate ARHI gene. Independent replication
321 is needed to confirm this lead. In addition, we found further evidence implicating *GRM7* in
322 ARHI.

323

324 **Acknowledgements**

325 The authors would like to express their most sincere gratitude to all the Saami volunteers who
326 have participated in this study. J.R.H. wishes to thank Gael Pressoir and Jianming Yu from
327 the Ed Buckler Lab for kindly providing help with their SAS macro. This work was funded by
328 the European Community (5th Framework project QLRT-2001-00331), by the University of
329 Antwerp (TOP project), by the Research Foundation - Flanders (FWO grant G.0163.09) and
330 by the State of Arizona. J.R.H. is a fellow of the Research Foundation - Flanders (FWO).

331

332 **Supplementary information** is available at the European Journal of Human Genetics'
333 website.

334

335 **References**

- 336 1. Davis A: Prevalence of Hearing Impairment; in: Davis A (ed): Hearing in adults.
337 London: Whurr Publishers Ltd, 1994, pp 43-321.
- 338 2. Gates GA, Mills JH: Presbycusis. *Lancet* 2005; **366**: 1111-1120.
- 339 3. Christensen K, Frederiksen H, Hoffman HJ: Genetic and environmental influences on
340 self-reported reduced hearing in the old and oldest old. *J Am Geriatr Soc* 2001; **49**:
341 1512-1517.
- 342 4. Gates GA, Couropmitree NN, Myers RH: Genetic associations in age-related hearing
343 thresholds. *Arch Otolaryngol Head Neck Surg* 1999; **125**: 654-659.
- 344 5. Karlsson KK, Harris JR, Svartengren M: Description and primary results from an
345 audiometric study of male twins. *Ear Hear* 1997; **18**: 114-120.
- 346 6. DeStefano AL, Gates GA, Heard-Costa N, Myers RH, Baldwin CT: Genomewide
347 linkage analysis to presbycusis in the Framingham Heart Study. *Arch Otolaryngol*
348 *Head Neck Surg* 2003; **129**: 285-289.

- 349 7. Garringer HJ, Pankratz ND, Nichols WC, Reed T: Hearing impairment susceptibility
350 in elderly men and the DFNA18 locus. *Arch Otolaryngol Head Neck Surg* 2006; **132**:
351 506-510.
- 352 8. Huyghe JR, Van Laer L, Hendrickx JJ *et al*: Genome-wide SNP-based linkage scan
353 identifies a locus on 8q24 for an age-related hearing impairment trait. *Am J Hum*
354 *Genet* 2008; **83**: 401-407.
- 355 9. Friedman RA, Van Laer L, Huentelman MJ *et al*: GRM7 variants confer susceptibility
356 to age-related hearing impairment. *Hum Mol Genet* 2009; **18**: 785-796.
- 357 10. Van Eyken E, Van Camp G, Van Laer L: The complexity of age-related hearing
358 impairment: contributing environmental and genetic factors. *Audiol Neurootol* 2007;
359 **12**: 345-358.
- 360 11. Van Laer L, Van Eyken E, Franssen E *et al*: The grainyhead like 2 gene (GRHL2),
361 alias TFCP2L3, is associated with age-related hearing impairment. *Hum Mol Genet*
362 2008; **17**: 159-169.
- 363 12. Bai U, Seidman MD, Hinojosa R, Quirk WS: Mitochondrial DNA deletions associated
364 with aging and possibly presbycusis: a human archival temporal bone study. *Am J*
365 *Otol* 1997; **18**: 449-453.
- 366 13. Fischel-Ghodsian N, Bykhovskaya Y, Taylor K *et al*: Temporal bone analysis of
367 patients with presbycusis reveals high frequency of mitochondrial mutations. *Hear Res*
368 1997; **110**: 147-154.
- 369 14. Cavalli-Sforza LL, Piazza A: Human genomic diversity in Europe: a summary of
370 recent research and prospects for the future. *Eur J Hum Genet* 1993; **1**: 3-18.
- 371 15. Lahermo P, Sajantila A, Sistonen P *et al*: The genetic relationship between the Finns
372 and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum*
373 *Genet* 1996; **58**: 1309-1322.

- 374 16. Lahermo P, Savontaus ML, Sistonen P *et al*: Y chromosomal polymorphisms reveal
375 founding lineages in the Finns and the Saami. *Eur J Hum Genet* 1999; **7**: 447-458.
- 376 17. Sajantila A, Lahermo P, Anttinen T *et al*: Genes and languages in Europe: an analysis
377 of mitochondrial lineages. *Genome Res* 1995; **5**: 42-52.
- 378 18. Tambets K, Rootsi S, Kivisild T *et al*: The western and eastern roots of the Saami--the
379 story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J*
380 *Hum Genet* 2004; **74**: 661-682.
- 381 19. Ingman M, Gyllensten U: A recent genetic link between Saami and the Volga-Ural
382 region of Russia. *Eur J Hum Genet* 2007; **15**: 115-120.
- 383 20. Johansson A, Vavruch-Nilsson V, Cox DR, Frazer KA, Gyllensten U: Evaluation of
384 the SNP tagging approach in an independent population sample--array-based SNP
385 discovery in Saami. *Hum Genet* 2007; **122**: 141-150.
- 386 21. Johansson A, Vavruch-Nilsson V, Edin-Liljegren A, Sjolander P, Gyllensten U:
387 Linkage disequilibrium between microsatellite markers in the Swedish Saami relative
388 to a worldwide selection of populations. *Hum Genet* 2005; **116**: 105-113.
- 389 22. Kaessmann H, Zollner S, Gustafsson AC *et al*: Extensive linkage disequilibrium in
390 small human populations in Eurasia. *Am J Hum Genet* 2002; **70**: 673-685.
- 391 23. Laan M, Paabo S: Demographic history and linkage disequilibrium in human
392 populations. *Nat Genet* 1997; **17**: 435-438.
- 393 24. Bonnen PE, Pe'er I, Plenge RM *et al*: Evaluating potential for whole-genome studies
394 in Kosrae, an isolated population in Micronesia. *Nat Genet* 2006; **38**: 214-217.
- 395 25. Service S, DeYoung J, Karayiorgou M *et al*: Magnitude and distribution of linkage
396 disequilibrium in population isolates and implications for genome-wide association
397 studies. *Nat Genet* 2006; **38**: 556-560.

- 398 26. Slatkin M: Linkage disequilibrium in growing and stable populations. *Genetics* 1994;
399 **137**: 331-336.
- 400 27. Terwilliger JD, Zollner S, Laan M, Paabo S: Mapping genes through the use of
401 linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations
402 with no demographic expansion. *Hum Hered* 1998; **48**: 138-154.
- 403 28. Van Eyken E, Van Laer L, Franssen E *et al*: KCNQ4: a gene for age-related hearing
404 impairment? *Hum Mutat* 2006; **27**: 1007-1016.
- 405 29. Johnson RA, Wichern DW: Applied Multivariate Statistical Analysis. New Jersey,
406 Prentice Hall, 2002.
- 407 30. Nelson EG, Hinojosa R: Presbycusis: a human temporal bone study of individuals
408 with downward sloping audiometric patterns of hearing loss and review of the
409 literature. *Laryngoscope* 2006; **116**: 1-12.
- 410 31. Schuknecht HF, Gacek MR: Cochlear pathology in presbycusis. *Ann Otol Rhinol*
411 *Laryngol* 1993; **102**: 1-16.
- 412 32. Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome
413 association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559-
414 575.
- 415 33. Yu J, Pressoir G, Briggs WH *et al*: A unified mixed-model method for association
416 mapping that accounts for multiple levels of relatedness. *Nat Genet* 2006; **38**: 203-
417 208.
- 418 34. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000
419 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661-
420 678.

- 421 35. Nelson EG, Hinojosa R: Presbycusis: a human temporal bone study of individuals
422 with flat audiometric patterns of hearing loss using a new method to quantify stria
423 vascularis volume. *Laryngoscope* 2003; **113**: 1672-1686.
- 424 36. Robertson NG, Khetarpal U, Gutierrez-Espeleta GA, Bieber FR, Morton CC: Isolation
425 of novel and known genes from a human fetal cochlear cDNA library using
426 subtractive hybridization and differential screening. *Genomics* 1994; **23**: 42-50.
- 427 37. Yamashiro S, Abe H, Mabuchi I: IQGAP2 is required for the cadherin-mediated cell-
428 to-cell adhesion in *Xenopus laevis* embryos. *Dev Biol* 2007; **308**: 485-493.
- 429 38. Kazmierczak P, Sakaguchi H, Tokita J *et al*: Cadherin 23 and protocadherin 15
430 interact to form tip-link filaments in sensory hair cells. *Nature* 2007; **449**: 87-91.
- 431 39. Bolz H, von Brederlow B, Ramirez A *et al*: Mutation of CDH23, encoding a new
432 member of the cadherin gene family, causes Usher syndrome type 1D. *Nat Genet*
433 2001; **27**: 108-112.
- 434 40. Bork JM, Peters LM, Riazuddin S *et al*: Usher syndrome 1D and nonsyndromic
435 autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel
436 cadherin-like gene CDH23. *Am J Hum Genet* 2001; **68**: 26-37.
- 437 41. Noben-Trauth K, Zheng QY, Johnson KR: Association of cadherin 23 with polygenic
438 inheritance and genetic modification of sensorineural hearing loss. *Nat Genet* 2003;
439 **35**: 21-23.
- 440 42. Van Eyken E, Van Camp G, Fransen E *et al*: The N-acetyltransferase 2 polymorphism
441 NAT2*6A is a causative factor for Age Related Hearing Impairment. *J Med Genet*
442 2007; **44**: 570-578.
- 443 43. Purcell S, Cherny SS, Sham PC: Genetic Power Calculator: design of linkage and
444 association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149-
445 150.

- 446 44. Weedon MN, Lango H, Lindgren CM *et al*: Genome-wide association analysis
447 identifies 20 loci that influence adult height. *Nat Genet* 2008; **40**: 575-583.
- 448 45. Klein RJ, Zeiss C, Chew EY *et al*: Complement factor H polymorphism in age-related
449 macular degeneration. *Science* 2005; **308**: 385-389.
- 450 46. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal
451 components analysis corrects for stratification in genome-wide association studies.
452 *Nat Genet* 2006; **38**: 904-909.
- 453 47. Patterson N, Price AL, Reich D: Population structure and eigenanalysis. *PLoS Genet*
454 2006; **2**: e190.
- 455
456

457 **Titles and Legends to Figures**

458 **Figure 1** Comparison of typical audiograms between subjects with high and low values for
459 PC1, PC2, and PC3. For each of the three traits, three subjects from the lower and upper 10%
460 extremes of the distribution of the trait are shown. Audiograms for both ears are given, to
461 illustrate the extent of intrasubject variability. Left and right ears correspond to filled and
462 open circles, respectively. For PC1, on average, hearing loss is much more pronounced in
463 individuals from higher extremes (top row) compared to the lower extremes (second row from
464 the top). Subjects with low values for PC2 (fourth row from the top) tend to have a much
465 more sloping audiogram, compared to subjects with high values for PC2 (third row from the
466 top). Subjects with high values for PC3 (second row from the bottom) tend to have a more
467 concave audiogram, compared to subjects in the lower extreme for PC3 (bottom row). Note
468 that the displayed audiometric data have not been adjusted for sex and age.

469

470 **Legends to Supplementary Figures**

471 **Supplementary Figure 1** Boxplots showing the effects of age and sex on hearing thresholds.
472 This figure illustrates that hearing deteriorates with age and that this deterioration is most
473 pronounced for the higher frequencies and in males. Note also that for the higher frequencies,
474 within a certain age class, the hearing thresholds tend to be more variable for males compared
475 to females. This is most apparent for 3 kHz and 4 kHz.

476

477 **Supplementary Figure 2** Evaluation of the mixed model versus the naive analysis. Eight
478 thousand randomly selected SNPs (approximately 10% of the SNPs meeting quality control
479 standards) were analysed with both methods and the results were compared for the three
480 principal components under study (PC1, PC2 and PC3 in respectively A, B and C). The
481 extreme left graph of each panel shows a quantile-quantile plot for the test statistics from the

482 mixed model analysis together with the genomic control inflation factor λ . The blue dashed
483 line goes through the first and third quartile. There is little evidence for residual confounding.
484 Scatter plots, together with Pearson correlation coefficients, are given in the middle graph of
485 each panel. The extreme right graph of each panel shows that the P-values from the naive
486 analysis tend to be too low relative to the more correct mixed model analysis.

487

488 **Supplementary Figure 3** Results of the statistical analysis for PC1 in the *GRM7* region. A)
489 The *GRM7* genomic region. The positions of the exons are indicated with numbered vertical
490 bars. The positions of the significant SNPs from our previous study⁹ and the sample sets in
491 which these results were obtained are indicated on top of the figure. The most significant
492 Saami *GRM7* SNP, rs161927, is also indicated on top. B) The individual significance values
493 for each SNP (expressed as $-\log_{10}P$ values) in the *GRM7* region are indicated with filled
494 diamonds on their respective locations within the *GRM7* locus. The nucleotide positions were
495 derived from the reference sequence with Genbank accession number NT_022517. C) The LD
496 structure within the *GRM7* locus for the general European population, as deduced from the
497 HapMap data (<http://www.hapmap.org/>). For this figure, only the SNPs that have been
498 analysed in the current study were selected in Haploview. D) The LD structure within the
499 *GRM7* locus for the Saami population based on our own genotyping data.

Table 1 Eigenvectors and eigenvalues for the first three principal components calculated from age and sex corrected hearing thresholds

<i>Principal Component (PC)</i>	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>
Eigenvalue	4.421	1.790	0.904
% of variation explained^a	0.491	0.199	0.100
Coefficients eigenvectors (loadings)^b			
res125	0.208	0.460	0.527
res250	0.288	0.470	0.287
res500	0.333	0.382	-0.186
res1000	0.348	0.186	-0.481
res2000	0.373	0.020	-0.408
res3000	0.393	-0.237	-0.109
res4000	0.358	-0.324	0.088
res6000	0.356	-0.328	0.276
res8000	0.305	-0.345	0.335

^aEigenvalue for the PC divided by the sum of all eigenvalues which corresponds to the total number of variables in the original dataset (9).

^bThe correlations between the PC and the age and sex corrected transformed hearing thresholds for each frequency can be obtained by multiplying each coefficient by the square root of its corresponding eigenvalue.

Table 2 Top 25 SNPs resulting from the mixed model analysis for Principal Component 1

<i>SNP</i>	<i>Chr.</i>	<i>Position</i>	<i>Alleles</i>	<i>Minor allele</i>	<i>MAF</i>	<i>Genotypes^a</i>	<i>P HWE^b</i>	<i>P Mixed</i>	<i>Gene Region^c</i>	<i>Pos. Rel. Gene^d</i>
rs1413042	13	38600755	C/T	C	0.228916	24/104/204	0.0432	2.91E-05	similar to RIKEN cDNA 8030451K01 (LOC387921)	downstream
rs1016111	6	14121677	C/G	C	0.367521	49/160/142	0.7308	5.86E-05	CD83 antigen precursor (Cell surface protein HB15)	upstream
rs1826882	15	19989036	A/C	C	0.113506	1/77/270	0.1024	9.41E-05	/	/
rs854273	1	69016062	G/T	T	0.247863	18/138/195	0.3898	0.000115	DEP domain containing 1	upstream
rs2825601	21	19766620	A/G	A	0.336182	39/158/154	0.9054	0.000134	/	/
rs1988145	5	92258480	A/C	A	0.228228	25/102/206	0.0195	0.000144	COUP transcription factor 1 (COUP-TF1)	upstream
rs9311746	3	60015052	C/T	T	0.381766	50/168/133	0.8221	0.000144	Bis(5'-adenosyl)-triphosphatase (EC 3.6.1.29)	intron
rs161927	3	7813242	A/G	G	0.065407	3/39/302	0.1663	0.000149	Metabotropic glutamate receptor 7 precursor (GRM7)	downstream
rs1358520	2	181148326	C/T	T	0.195157	9/119/223	0.1738	0.000187	Ubiquitin-conjugating enzyme E2 E3 (EC 6.3.2.19)	upstream
rs2148418	13	38564233	A/G	G	0.215714	20/111/219	0.2675	0.000194	similar to RIKEN cDNA 8030451K01 (LOC387921)	downstream
rs1661609	7	52720759	A/T	T	0.150000	7/91/252	0.8359	0.000208	/	/
rs572619	11	115738853	A/T	T	0.361823	45/164/142	0.9080	0.000211	BUD13 homolog	downstream
rs10510777	3	55212778	A/G	G	0.472934	76/180/95	0.6684	0.000227	calcium channel, voltage-dependent, alpha 2/delta 3 subunit	downstream
rs2182406	10	129456655	C/T	T	0.133903	8/78/265	0.4864	0.000231	transmembrane protein 12	downstream
rs9301169	13	106353066	C/T	C	0.250712	21/134/196	0.8870	0.000234	/	/
rs1388654	1	68985603	A/G	G	0.295714	26/155/169	0.3039	0.000249	DEP domain containing 1	upstream
rs3751591	15	49394002	A/G	G	0.148415	14/75/258	0.0100	0.000250	Cytochrome P450 19A1 (EC 1.14.14.1)	intron
rs7620394	3	55206368	C/T	C	0.304843	33/148/170	0.9003	0.000253	leucine-rich repeats and transmembrane domains 1	upstream
rs10503920	8	32548231	A/G	G	0.274929	34/125/192	0.0454	0.000259	Neuregulin-1, sensory and motor neuron-derived factor isoform	intron
rs10497334	2	168549782	A/G	A	0.081197	1/55/295	0.7146	0.000266	STE20/SPS1-related proline-alanine-rich protein kinase (EC 2.7.11.1)	intron
rs7357631	9	99568141	A/G	A	0.497151	87/175/89	1.0000	0.000282	Forkhead box protein E1 (Thyroid transcription factor 2)	upstream
rs2055439	5	75511014	A/G	A	0.403458	61/158/128	0.3166	0.000292	Synaptic vesicle glycoprotein 2C	intron
rs10515629	5	148875068	C/T	C	0.196581	15/108/228	0.6132	0.000303	Casein kinase I isoform alpha (EC 2.7.11.1)	intron
rs2037344	8	121088918	C/T	C	0.348665	39/157/141	0.7191	0.000339	DEP domain containing 6	intron
rs3912622	11	102626683	A/C	C	0.111748	8/62/279	0.0570	0.000355	DYNC2H1 protein	intron

^aThe number of homozygotes for the minor allele/ the number of heterozygotes/ the number of homozygotes for the major allele.

^bThe Hardy-Weinberg equilibrium P-value should be larger than 0.001 for the SNP genotyping to be valid.

^c"/" indicates that this SNP was not located within a gene region.

^d"upstream" or "downstream" was based on the Affymetrix annotation.

Chr.: chromosome; MAF: Minor Allele Frequency; P HWE: P-value resulting after Hardy-Weinberg equilibrium testing; P Mixed: P-value resulting after analysis with the mixed model (indicated in bold); Pos. Rel. Gene: Position of the SNP relative to the gene listed in the previous column

Table 3 Top 25 SNPs resulting from the mixed model analysis for Principal Component 2

<i>SNP</i>	<i>Chr.</i>	<i>Position</i>	<i>Alleles</i>	<i>Minor allele</i>	<i>MAF</i>	<i>Genotypes^a</i>	<i>P HWE^b</i>	<i>P mixed</i>	<i>Gene Region^c</i>	<i>Pos. Rel. Gene^d</i>
rs10499138	6	128533875	A/T	A	0.253561	30/118/203	0.0473	1.70E-06	Receptor-type tyrosine-prot phosphatase kappa precursor (EC 3.1.3.48)	intron
rs10491923	9	76139000	A/G	A	0.365672	49/147/139	0.3460	2.79E-05	Nuclear receptor ROR-beta (Nuclear receptor RZR-beta)	upstream
rs10499136	6	128419828	A/C	C	0.296296	36/136/179	0.2002	3.58E-05	Receptor-type tyrosine-prot phosphatase kappa precursor (EC 3.1.3.48)	intron
rs9308868	2	103727852	C/T	C	0.268895	25/135/184	1.0000	6.45E-05	U3 small nucleolar RNA	upstream
rs1555167	6	62465026	G/T	T	0.398860	57/166/128	0.8239	9.62E-05	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs9294311	6	62470112	C/T	T	0.398860	57/166/128	0.8239	9.62E-05	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs9294314	6	62470480	A/G	G	0.398860	57/166/128	0.8239	9.62E-05	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs10485162	6	128639684	C/T	T	0.145299	12/78/261	0.0530	0.000122	Receptor-type tyrosine-prot phosphatase kappa precursor (EC 3.1.3.48)	intron
rs514878	5	73705460	A/G	A	0.058405	2/37/312	0.3294	0.000132	Ectoderm-neural cortex 1 protein (ENC-1)	downstream
rs10520939	5	27970143	C/T	T	0.082813	5/43/272	0.0516	0.000149	U6 spliceosomal RNA	downstream
rs1201494	6	62547842	A/G	A	0.398571	56/167/127	0.9116	0.000165	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs304408	18	1095249	A/G	G	0.245588	28/111/201	0.0288	0.000175	Putative C18orf2 variant 3	downstream
rs768166	10	8087556	A/C	C	0.228198	18/121/205	1.0000	0.000175	Transcription initiation factor TFIID subunit 3	intron
rs304409	18	1096342	A/G	A	0.246356	25/119/199	0.2439	0.000180	Pituitary adenylate cyclase-activating polypeptide precursor (PACAP)	downstream
rs4936417	11	117409046	A/G	G	0.200000	20/100/230	0.0460	0.000202	Eukaryotic type signal recognition particle RNA	upstream
rs4386936	8	135859109	A/C	C	0.219373	15/124/212	0.6414	0.000221	Zinc finger prot 406 (Protein ZFAT)	upstream
rs211004	5	161577376	C/T	T	0.370640	53/149/142	0.2033	0.000222	Gamma-aminobutyric-acid receptor gamma-2 subunit precursor	downstream
rs522914	6	62592457	A/T	A	0.400285	57/167/127	0.9115	0.000224	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs565795	6	62597708	A/G	G	0.400285	57/167/127	0.9115	0.000224	KH dom-containing RNA-binding signal transduction-associated prot 2	intron
rs265360	6	129579569	A/G	A	0.272727	33/120/188	0.0404	0.000247	Laminin alpha-2 chain precursor (Laminin M chain)	intron
rs9312176	4	67882411	A/G	A	0.456522	74/167/104	0.6647	0.000254	similar to Interferon-induced transmembrane prot 3	downstream
rs1544044	20	14459168	C/T	T	0.465318	72/178/96	0.5892	0.000257	OTTHUMP00000030317	intron
rs10493937	1	101115238	A/T	A	0.472934	85/162/104	0.1650	0.000267	Exostosin-like 2 (EC 2.4.1.223)	intron
rs1358815	6	57471407	G/T	G	0.464183	73/178/98	0.6680	0.000274	DNA primase large subunit (EC 2.7.7.-)	intron
rs6489314	12	1627431	A/C	A	0.341642	47/139/155	0.0916	0.000300	Adiponectin receptor prot 2	upstream

^aThe number of homozygotes for the minor allele/ the number of heterozygotes/ the number of homozygotes for the major allele.

^bThe Hardy-Weinberg equilibrium P-value should be larger than 0.001 for the SNP genotyping to be valid.

^c“/” indicates that this SNP was not located within a gene region.

^d“upstream” or “downstream” was based on the Affymetrix annotation.

Chr.: chromosome; MAF: Minor Allele Frequency; P HWE: P-value resulting after Hardy-Weinberg equilibrium testing; P Mixed: P-value resulting after analysis with the mixed model (indicated in bold); Pos. Rel. Gene: Position of the SNP relative to the gene listed in the previous column; Dom: domain; Prot: protein

Table 4 Top 25 SNPs resulting from the mixed model analysis for Principal Component 3

<i>SNP</i>	<i>Chr.</i>	<i>Position</i>	<i>Alleles</i>	<i>Minor allele</i>	<i>MAF</i>	<i>Genotypes^a</i>	<i>P HWE^b</i>	<i>P Mixed</i>	<i>Gene Region^c</i>	<i>Pos. Rel. Gene^d</i>
rs457717	5	75956728	C/T	T	0.251429	26/124/200	0.2589	3.55E-07	Ras GTPase-activating-like prot IQGAP2	intron
rs1697845	5	75958260	C/T	C	0.293447	37/132/182	0.0935	1.63E-05	Ras GTPase-activating-like prot IQGAP2	intron
rs989414	2	151552224	C/T	T	0.061960	3/37/307	0.1326	3.72E-05	CDNA FLJ45645 fis, clone CTONG2003517	downstream
rs9310194	3	70713297	A/T	A	0.334770	35/163/150	0.3994	4.96E-05	Microphthalmia-associated transcription factor	downstream
rs10514104	5	76843891	A/G	G	0.113506	4/71/273	1.0000	5.08E-05	/	/
rs7913244	10	10224371	A/T	T	0.102857	5/62/283	0.3902	5.69E-05	/	/
rs2299641	11	17397566	C/G	G	0.200000	14/112/224	1.0000	7.90E-05	ATP-binding cassette transporter sub-family C member 8	intron
rs10494078	1	108157301	C/T	T	0.472303	72/180/91	0.3860	9.16E-05	Prot vav-3	intron
rs4636135	7	103406504	A/G	A	0.246418	19/134/196	0.6650	0.000100	Reelin prec (EC 3.4.21.-)	intron
rs6539455	12	107744032	C/G	G	0.115607	9/62/275	0.0310	0.000101	Prot phosphatase Slingshot homolog 1 (EC 3.1.3.48)	intron
rs10497072	2	151581781	A/G	A	0.058405	1/39/311	1.0000	0.000135	Rho-related GTP-binding prot RhoE (Rho family GTPase 3)	upstream
rs4664883	2	151583181	A/C	A	0.058405	1/39/311	1.0000	0.000135	CDNA FLJ45645 fis, clone CTONG2003517	downstream
rs1500237	5	98643065	C/T	T	0.205128	17/110/224	0.5116	0.000145	Y RNA	upstream
rs6433104	2	151559112	C/T	T	0.058908	1/39/308	1.0000	0.000151	Rho-related GTP-binding prot RhoE (Rho family GTPase 3)	upstream
rs10483426	14	32670589	C/T	T	0.452991	79/160/112	0.1326	0.000163	Neuronal PAS domain-containing prot 3 (Neuronal PAS3)	intron
rs1940484	11	98447082	C/T	C	0.475714	82/169/99	0.5920	0.000164	/	/
rs9301099	13	105483574	A/G	G	0.169516	14/91/246	0.1321	0.000183	/	/
rs2527756	8	5390145	C/T	C	0.188571	10/112/228	0.4856	0.000199	/	/
rs10512077	9	80741480	A/G	A	0.417379	67/159/125	0.2269	0.000201	/	/
rs42064	7	95557193	A/T	A	0.319088	42/140/169	0.1399	0.000210	Cytoplasmic dynein 1 intermediate chain 1	intron
rs1363556	5	144886649	C/T	T	0.498457	90/143/91	0.0351	0.000216	7SK RNA	upstream
rs1260764	1	64316010	G/T	G	0.176638	12/100/239	0.7133	0.000221	Tyrosine-prot kinase transmembrane receptor ROR1 prec (EC 2.7.10.1)	intron
rs1260772	1	64313964	A/G	G	0.176638	12/100/239	0.7133	0.000221	Tyrosine-prot kinase transmembrane receptor ROR1 prec (EC 2.7.10.1)	intron
rs1559776	15	55746920	C/T	T	0.150997	7/92/252	0.8356	0.000236	GRINL1A combined prot isoform 1	intron
rs4809590	20	44518234	C/T	T	0.102564	7/58/286	0.0732	0.000253	/	/

^aThe number of homozygotes for the minor allele/ the number of heterozygotes/ the number of homozygotes for the major allele.

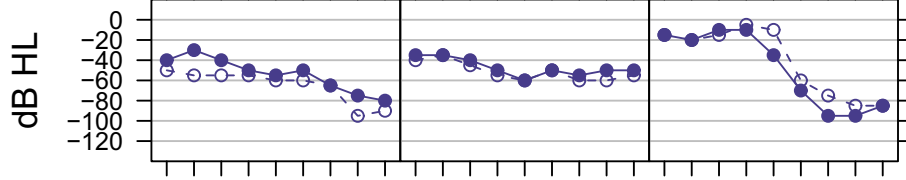
^bThe Hardy-Weinberg equilibrium P-value should be larger than 0.001 for the SNP genotyping to be valid.

^c"/" indicates that this SNP was not located within a gene region.

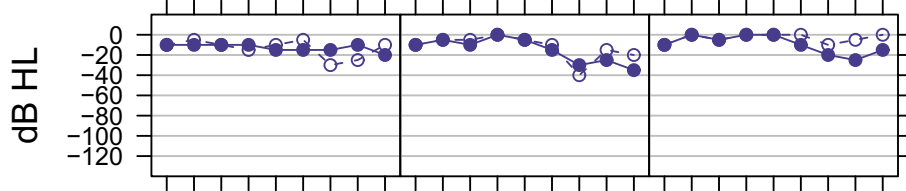
^d"upstream" or "downstream" was based on the Affymetrix annotation.

P-values are indicated in bold. Chr.: chromosome; MAF: Minor Allele Frequency; P HWE: P-value resulting after Hardy-Weinberg equilibrium testing; P Mixed: P-value resulting after analysis with the mixed model (indicated in bold); Pos. Rel. Gene: Position of the SNP relative to the gene listed in the previous column; Prot: protein; Prec: Precursor

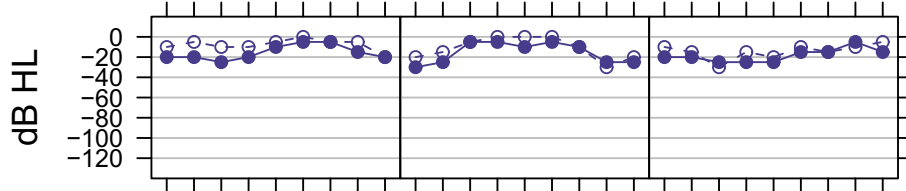
Higher extreme PC1



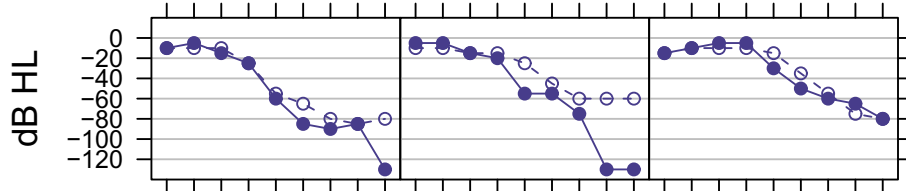
Lower extreme PC1



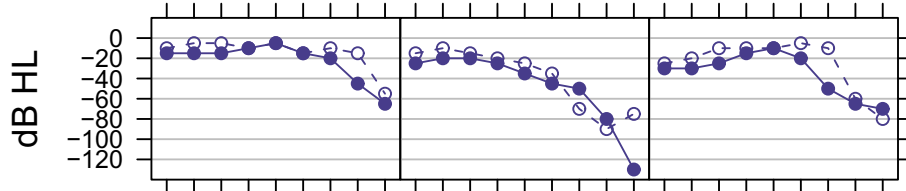
Higher extreme PC2



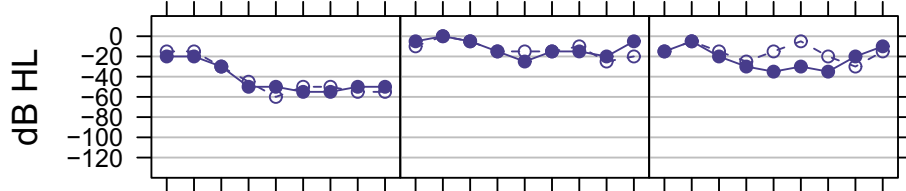
Lower extreme PC2



Higher extreme PC3



Lower extreme PC3



Frequency (kHz)