ORIGINAL RESEARCH

A novel high-throughput analysis approach: immune response-related genes are upregulated in age-related hearing loss

Yolande Tra[1]
Robert D Frisina[2]
Mary D'Souza[3]

[1]MD Poison Center, School of Pharmacy, University of Maryland, Baltimore, MD, [2]Department of Chemical and Biomedical Engineering, College of Engineering, University of South Florida, Tampa FL, [3]Department of Surgery, and the Wilmot Cancer Center, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

Correspondence: Robert D Frisina
Department of Chemical and Biomedical Engineering, 4202 E Fowler Avenue, ENB 118, University of South Florida, Tampa FL 33620, USA
Tel +1 813 974 4394
Fax +1 813 974 3651
Email rfrisina@usf.edu

Background: Presbycusis is defined as the hearing loss that occurs with aging. Genes that are responsible for this clinical condition have still not been identified. The present study explores gene discovery for age-related hearing loss using the CBA/CaJ mouse model of age-related hearing loss.

Methods and results: In the present investigation, in addition to a gene-based analysis, a molecular pathway analysis was performed to evaluate differences in gene expression in 15 aged mice with age-related hearing loss, and 25 young to middle-aged mice with normal hearing. Using three different statistical approaches, the gene-based analysis revealed four common probe sets, ie, *Ctss* (Cathespin), *Csnk* (casein kappa), *Mpeg1* (macrophage-expressed gene 1), and *Clecsf12* (C-type calcium-dependent carbohydrate recognition domain lectin, superfamily member 12) that are significantly upregulated with age and hearing loss in the mouse cochlea (inner ear). Relative real-time polymerase chain reaction was used for quantitative, sensitivity analysis of gene transcription changes and confirmed the gene microarray results. Three strategies for pathway analysis were consistent with these gene expression discoveries, and revealed a common immune response pathway. More specifically, the analysis suggests that B cell-mediated humoral immune function plays an important role in the underlying etiology of presbycusis, similar to certain other neurodegenerative diseases. Other top pathways from this analysis included those involving dendritic antigen-presenting cells, carbohydrate binding, G-protein coupled receptor binding, and epithelial-to-mesenchymal transition pathways.

Conclusion: Combining microarray gene discovery, polymerase chain reaction, and multiple pathway analyses revealed insights into immune system involvement in the progression of age-related hearing loss.

Keywords: high-throughput gene analysis, gene microarray, presbycusis, bioinformatics, biostatistics, auditory system

Introduction

Presbycusis, or hearing loss due to aging, involves genetic, cellular, and systems-level changes in the auditory part of the inner ear, ie, the cochlea, and the parts of the brain used for hearing, ie, the central auditory system.[1,2] Presbycusis occurs in all mammalian species, including humans, and is the most prevalent communication disorder and one of the top three most common chronic medical conditions of the aged.

However, the mechanisms underlying changes in gene expression due to aging in the cochlea remain unresolved. Several auditory studies have examined the effects of altering levels of specific molecules, including neurotransmitters and their agonists and antagonists, aiming at understanding the synaptic changes that accompany age-related hearing loss. Indeed, some of the physiological effects of gamma amino butyric

107

acid blockade activity, particularly in the central auditory system, resemble the negative effects of aging on auditory processing.[3–8] However, the structural and functional aspects of hearing loss due to age (over a longer period) are likely to be mediated by transcriptional changes in a complex set of genes, including some from those used to form the organ of Corti during development, or related to other neurodegenerative diseases of old age. The identity of such genes and how they are regulated with age is currently unknown for the aging auditory system.

Although some of the genes in the cochlea and central auditory system associated with critical developmental periods have been discovered,[9] there are only a few studies identifying sets of genes that are upregulated or downregulated in the aging auditory system.[10–13] Using a novel multivariate search, Xiao et al identified 10 probe sets exhibiting changes with age, of which six were functionally related to the immune system, three to post-translational protein modifications, and one to mammary gland function.[14]

In the present report, DNA microarrays were utilized to examine large-scale changes in gene expression in the aging cochlea of the CBA/CaJ mouse, using quantitative analyses of single genes as well as computational analyses of gene sets. CBA mice have proved to be a very useful model for studying age-related hearing loss, including a recent investigation where it was shown that mice combine pathology of hair cells, neurons, and stria with associated endocochlear potential decline.[15] The latter two are more pronounced, and in females these cochlear pathologies accelerate after menopause. Hearing loss was measured by auditory brainstem response (ABR) thresholds and distortion-product otoacoustic emission (DPOAE) levels. These physiological measurements of hearing, age, and gender were used as covariates to assess the usefulness of gene expression in a microarray-based discovery approach. It was of interest to identify features (gene expression changes) that characterize presbycusis, but also to know their additional predictive value for key characteristics of the phenotype. We explored the effects of adjusting for functional hearing measurement on the association between hearing status and gene expression. The data analysis tool, partial least squares (PLS),[16–18] combined with random forest (RF) procedures, were used to determine the effect of variations in gene expression based on top PLS components.

PLS is a useful technique when the number of predictors (probe sets on the gene microarray, about 22,000) is much larger than the number of observations (samples, number of mice, or replicates of 40). Compared with the classical least squares methodology, PLS is an appropriate method when predictors such as probe sets are highly correlated. It summarizes gene expression via dimension reduction. The outcome is a small number of factors (usually two), which is a linear combination of the original predictors that have maximum correlation with the dependent variable (hearing status).

RF is an artificial intelligence method that has the same listed property as PLS. It is a valuable technique to rank genes by taking into account the interaction effect among genes in the genome. Simulation studies show that when more interactions are present in the data, RF outperforms a univariate method, does not overfit the data, and does not rely on fine tuning parameters to achieve excellent performance for a particular experiment. It returns several measures of variable importance. For the present study, the average reduction in the accuracy of the class votes was selected.[19] RF has shown performance comparable with other classification methods such as the support vector machine and many neural networks, and identifies the smallest possible set of variables (excluding redundant variables) that achieves good predictive performance. In the present investigation, RF determined which genes are likely to be important predictors of hearing loss. A combination of the PLS and RF approaches provides an assessment of the usefulness of gene expression in a biological classification approach. However, for practicality and specificity, we additionally applied RF alone to search for genes/functional hearing measurements that distinguish between subject groups with good hearing and with hearing loss.

Methods
Subjects
Animal subjects and laboratory procedure details were the same as our previous microarray investigations.[11,12] CBA/CaJ mice were bred in-house according to the University of Rochester Animal Use Committee protocols. Original breeding pairs were obtained from Jackson Laboratories (Bar Harbor, ME). All animals had similar environmental and non-ototoxic histories, being raised together in a relatively quiet, dedicated vivarium room. All mice had their ear canals and tympanic membranes (very thin/transparent in the mouse) inspected with an operating microscope prior to inclusion in the study, and all were normal.

Functional hearing assessment
As reported previously,[11] the subjects were classified into four groups based on age and hearing ability, ie, young controls with good hearing, middle-aged mice with good hearing, old-aged

mice with mild presbycusis, and old-aged mice with severe presbycusis. In this study, the main interest was not in single genes but rather in a group of genes associated with a pathway or having a common location in the genome, so a differential gene expression between different conditions may be better performed using a global test to avoid multiple testing problems. Global analysis of covariance (Bioconductor Package) was used for global testing of differential gene expression between the groups. To avoid alpha inflation due to correlated data and the effects of non-normality, tests for significance of the resulting F ratios were performed using permutations. This suggested that the overall gene expression profile for all genes was associated ($P = 0.0215$) with hearing status using two groupings, ie, middle-aged mice with good hearing and young controls, and old-aged mice with mild presbycusis and severe presbycusis. Thus, in the statistical analysis, two groups of mice were used, with DPOAE amplitudes and ABR threshold measurements being similar for the mice in each group. Table 1 shows the age of the mice at the time of testing, the averages and standard deviations of ABR threshold, and DPOAE levels for each group at representative frequencies. As for most aspects of aging, the results show a larger spread of hearing measurements for the presbycusis group. There were 25 mice with good hearing (56% female) and 15 mice with presbycusis (57% female).

Cochlea isolation

Tissue isolation, RNA preparation, cDNA synthesis, GeneChip target hybridization, washing, staining, and scanning were performed in the same way as in our previous reports.[10,12,25] Microarrays (MOE430A) were purchased from Affymetrix (Santa Clara, CA), and each chip contained 22,690 unique probe sets (information available at http://www.affymetrix.com). Oligonucleotide identifiers were

25 nucleotides in length, and centrally positioned with a nine probe set of perfect match and mismatch rows designed with *Mus musculus* (mouse) RefSeq sequences. Each array had 14,484 full-length transcripts, 9450 transcripts of non-expressed sequence tags (excluding full length) and 21,103 transcripts from expressed sequence tags with strong evidence of polyadenylation. The sequence selection region was a 600 base pair area selected from the consensus, with regions based on strong evidence for polyadenylation, a full-length 3' end, and consensus sequence ends (locus link database http://www.affymetrix.com).

Normalization of microarray data

Hybridization intensities for each chip were scaled in comparison with young adult controls (eight chips). Intensity values were normalized using the robust multichip average.[26] This strategy fits a robust linear model at the probe level and tends to include low intensity perfect match signals relative to mismatch signals.

RNA and cDNA synthesis for relative real-time PCR

As in our previous published reports, total RNA was obtained from samples retained from the original microarray experiments.[10–13] Using the NuGen kit Protocol according to the manufacturer's instructions (http://www.nugeninc.com/nugen/), 5 ng was reverse transcribed to cDNA.

Taqman primer/probes for relative real-time PCR

Taqman primer/probes (ABI, Applied Biosystems Inc, Framingham, MA) were used to confirm the microarray data. The probes were tagged with a fluorescence dye (FAM) for passive fluorescence (ROX). Relative real-time PCR analysis of samples from the GeneChip experiments, as described by Livak and Schmittgen,[22] were done on triplicates of each sample on an ABI 7500 thermocycler (PE, Applied Biosystems) with a TaqMan universal master mix (ABI). The following PCR profile was applied: incubation for 2 minutes at 50°C, then initial denaturation at 10 minutes at 95°C, followed by 50 cycles of denaturation for 15 seconds at 95°C, and 1 minute of annealing at 60°C.

Statistical analysis
Analysis of microarray data combined with functional hearing measurements

Two groups were used for the analysis based on the significant association between the gene expression profile for all genes

Table 1 Age and functional hearing measurement means and standard deviations for two CBA mouse subject groups

	Good hearing	Presbycusis
Age at ABR	9.4 (4.27)	28.93 (3.33)
3 kHz ABR	60.4 (10.79)	81.33 (18.07)
6 kHz ABR	31 (9.46)	64.33 (24.78)
12 kHz ABR	13.2 (5.93)	42 (18.69)
24 kHz ABR	19.4 (7.68)	53.67 (22.16)
32 kHz ABR	27.4 (7.92)	64.33 (22.11)
48 kHz ABR	32 (7.5)	70 (17.83)
Age at DPOAE	9.46 (4.34)	29.23 (3.23)
Low frequency DPOAE (5–14 kHz)	18.13 (4.62)	−0.83 (12.49)
Mid frequency DPOAE (15–29 kHz)	17.93 (4.37)	−5.3 (13.19)
High frequency DPOAE (31–45 kHz)	9.09 (5.82)	−9.66 (11.52)

Abbreviations: ABR, auditory brainstem response; DPOAE, distortion-product otoacoustic emission.

and the two physiological states of hearing. A combination of PLS and RF was performed to assess the usefulness of gene expression over functional hearing measurements. In the context of classification, PLS is a multivariate data analysis technique that reduces the dimensions of the predictors. It is known to have a good prediction accuracy with high-dimensional microarray data.[16–18] It extracted two components, ie, a linear combination of probes. The RF variable importance measure then ranks the two components and the hearing measurements. To identify the important probe sets in these linear combinations, differentiating gene hearing status, RF alone was performed using the variable importance measure output. This algorithm contains a cross-validation step which states a priori that there is no need to divide the microarray data into two sets, ie, a training set and a test set. However, for fair evaluation of the classifier, at each cross-validation run, samples of mice were randomly split into train and test sets, and assigned a class label (hearing status) within each set, using two splitting ratios. Two usual choices of size ratios between the two sets were utilized, ie, 2:1 and 4:1.[17] For example, in a ratio of 2:1, one third of the samples were randomly selected for a test set with two class labels. With bootstrap sampling, the indices of samples included in each set were randomly selected for replacement with specific probabilities corresponding to the split ratio. One third of the original data were left out of the sample when the ith tree was constructed. These out-of-bag (OOB) data were used to obtain an unbiased estimate of the classification error for the ith tree. The error rate was estimated using cross-validation of the OOB samples in the test set that was not used for constructing the classifier. We selected the mean decrease in accuracy measure to rank the importance of variables (probe sets and hearing measurements) in terms of prediction. This measure determines how variable importance contributes to prediction accuracy. The algorithm works as follows: for each tree, the values of the mth variable are randomly rearranged for the out-of-bag set; it then puts this permuted set further down the tree, and gets new classifications for the forest. The mean decrease in accuracy is the difference between the OOB error rate for the randomly permuted mth variable and the original OOB error rate.

Differentially expressed genes

To detect differences in gene expression between mice with good hearing and mice with hearing loss, ie, presbycusis, the moderated t statistic from the Limma was used.[23] This method combines classical linear models with Bayesian inference for assessing statistical significance. The moderated t statistic

makes a more stable estimate of gene-specific variance by using information about variability across all genes. This technique avoids the problem of artificially reduced variance estimates for genes that show high reproducibility by chance alone. The Bayesian approach to variance estimation allows one to analyze unbalanced experimental designs and small sample size with Limma. A false discovery rate equal to 1% for multiple testing corrections was used following the Benjamini and Hochberg method.[24] Probes were ranked by their adjusted P values in ascending order.

Pathway analysis

Not all genes have been annotated in currently available databases or pathway annotation lists. Pathway names and constituents are not always the same, and sometimes overlap. Moreover, there are genes in the microarray of the present study that are not yet classified in a pathway. Thus, we considered three pathway analysis tools for comparison and consistency of results. These were: tool 1, a pathway-based classification using random forests; tool 2, GOstats, a bioinformatics tool for gene ontology term association, developed in Bioconductor for the Gene Ontology Consortium database;[25] and tool 3, the SigPathway package.[26] For tool 1, the pathways are from gene ontology, KEGG, BioCarta, GenMAPP, and manually curated sources. Important ranked pathways are identified through the OOB estimates of classification error. The default cross-validation in the RF algorithm assesses the error rate.

An advantage of tool 2 (GOstats) is that it allows users to select a specific level of abstraction or biological process. Tool 3 is an extension of the gene set enrichment method.[27,28] It has been shown to be very efficient for finding powerful gene sets for contrasting phenotypes. The pathway annotations are curated from gene ontology, KEGG, BioCarta, BioCyc, and SuperArray. The analysis has two objectives, ie, to test if a gene set shows a significant pattern of association with the phenotype compared with the rest of the genes, and to test if a gene set contains a high proportion of genes with expression levels associated with the phenotype of interest. Differential expression of each gene set was assessed using two test statistics, ie, near-threatened and not evaluated, defined as follows. Near-threatened (where gene labels are permuted) measures the strength of association with hearing status of a given gene set compared with other gene sets. For each gene set, a composite t score was obtained. Not evaluated (where sample labels are permuted) measures the strength of association with hearing status for genes within a gene set. The modified gene set enrichment method program

ranks the gene sets based on average rank-order of near-threatened and not evaluated values. Given that the goal is to minimize false positives, both rank orders are required to be high. To adjust for multiple testing, the output also includes a statistical significance for each gene set with a q value (false discovery rate). The first two tools employ the whole genome and do not need any gene preselection.

Promoter analysis

Promoter analysis was done with Genomatrix software (v 2.3; Genomatix Software GmbH, Munich, Germany) using the top four genes discovered in the present Gene-Chip study, namely *Ctss*, *Csnk*, *Mpeg1*, and *Clecsf12* in *M. musculus* (mouse). A number of conserved transcription binding sites were predicted, using the Genomatix tool, DiAlign. The DiAlign transcription factor task of the GEMS Launcher was used to check for conserved transcription factor binding sites from six species, ie, *Homo sapiens, Pan troglodytes, Macaca mulatta, Canis familiaris, Bos taurus, Rattus norvegicus* (Matrix Family Library, Version 8.0, Vertebrates; Genomatix: 690 matrices from 162 families), and the predicted transcription factor module was MYOD_VTBP_01 in the *Ctss* gene and transcription factor module MYOD_SRFF_01 in the *Csnk* gene of the present investigation. The promoter is conserved in six species (*H. sapiens, P. troglodytes, M. mulatta, C. familiaris, B. taurus,* and *R. norvegicus*), and this promoter codes for 1–6 relevant transcripts in the *Ctss, Csnk, Mpeg1,* and *Clecsf12* genes.

Results
Statistical analysis
Changes in gene expression in aging mice

Global analysis of covariance indicated that the overall gene expression profile is associated with hearing status using a binary grouping (hearing versus nonhearing, $P = 0.0215$). The comparison with two groups yielded 89 differentially expressed genes ranked by B statistics at a false discovery rate of 1%. The moderated t statistic, computed in a linear model for microarray (Limma), assessed statistical significance for gene expression data. It represents the log odds that the gene is differentially expressed between the two groups. A high moderated t statistic indicates strong evidence of differential expression. For each gene, a log 2 fold change, a point estimate of the differential expression, was also obtained. The present microarray data analysis confirmed the identification of four probe sets as upregulated with age and hearing loss, ie, *Ctss* (Cathespin), *Csnk* (casein kappa), *Mpeg1* (macrophage-expressed gene 1), and *Clecsf12* (C-type

calcium-dependent carbohydrate recognition domain lectin, superfamily member 12). These probes form a set of overlapping probes from previous approaches (Figure 1).

Cross validation of the class prediction

The following were available at the time of testing: age in months for ABR (ABRAge) and DPOAE (DPOAEAge), average ABR thresholds for sound frequencies covering the mouse range of hearing (3, 6, 12, 24, 32, and 48 kHz), and DPOAE levels for representative frequency bands appropriate for mouse audition (low, middle, and high, see Table 1). These predictors were used as covariates when PLS and RF combined was run with different seeds (PLS optimized with two components). Each optimal component represented a linear combination of probes. The error rates for the test set were 0%. The first PLS component (the most informative one) was ranked as high as the covariates that were also assessed as good predictors (error rate = 0%) of hearing status in a model only using covariates. Results showed that the variable importance rank depends on the basis of splitting. Using the ratio 2:1, the first component was ranked as second in importance to DPOAEAge. Four probes were ranked as important as age. For the ratio 4:1, the first component ranked below ABRAge and DPOAEAge, but was more important than the other covariates. This result showed that gene expression was a potential significant predictor of hearing status. To identify the specific genes competing with the covariates, the RF method was run alone. There were 22 females and 18 males in the entire study. Gender was found to be a weak predictor. When several runs with the same parameter settings were performed with different seed values, with only the covariates, the error rate was 2.63%. The analyses revealed the following ranking of subject characteristics (from most to least important): DPOAEAge 32 kHz; ABRAge 48 kHz, 12 kHz, 24 kHz, 6 kHz, and 3 kHz; and gender.

The following scenarios for predictors were considered in two different models, ie, a model with only probe sets (Model 1) and a model with probe sets, age, gender, and hearing measurements (Model 2). In Model 1, the classification gave satisfactory results (with a small error rate) showing that the microarray data exhibited a good relationship with hearing status. The outcome from Model 2 suggested that the microarray data were weaker than other subject data as a predictor of hearing status, but not redundant. There were several gene probes that stood out as being as important as the phenotypic hearing measurements.

For a better estimate of the error rate, labels were permuted 10 times, giving different training and test sets for each splitting
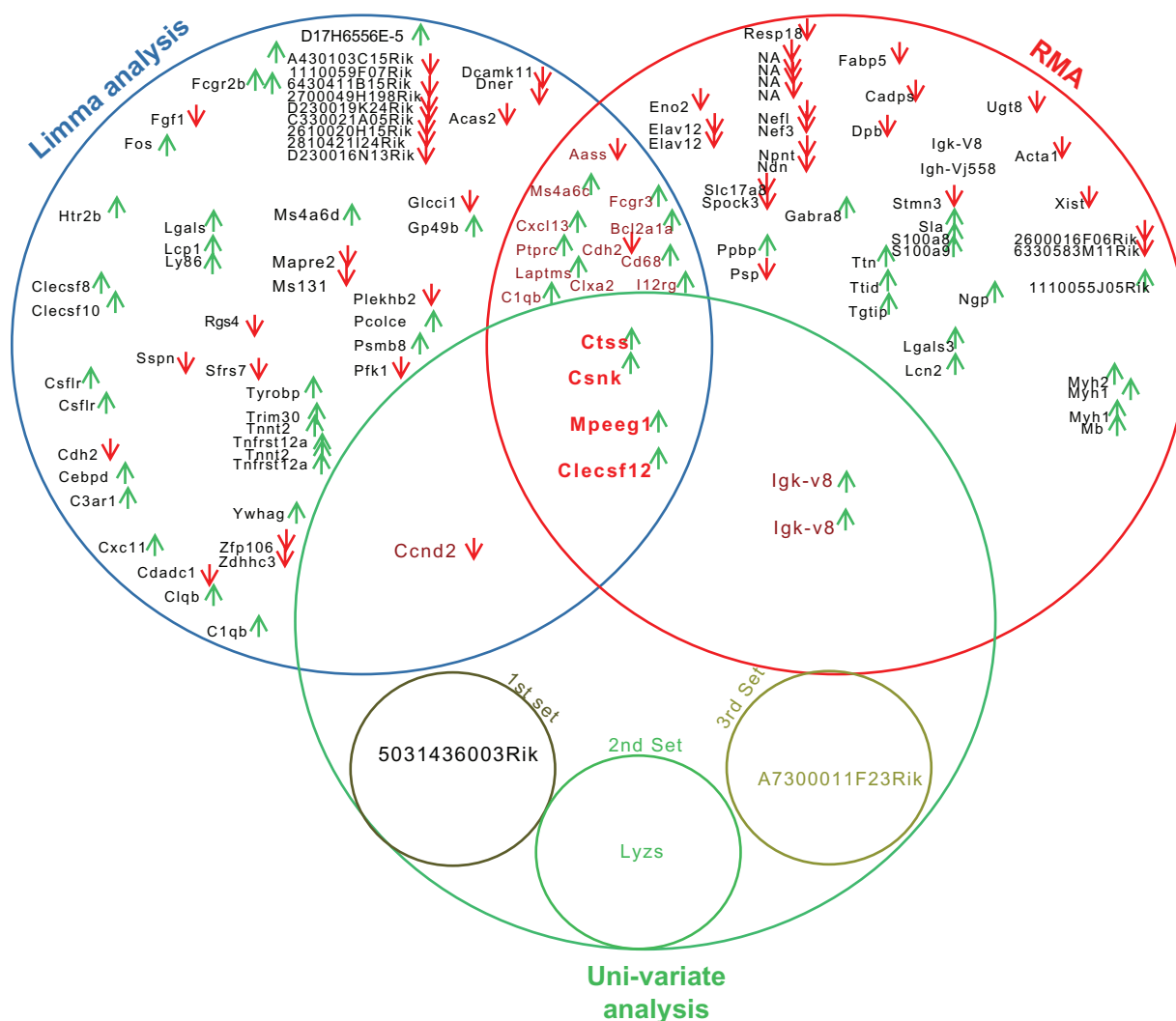
Figure 1 Venn-diagram: Uni-variate analysis (Green), RMA (Red) – RMA normalized gene expressions subjected to one-way ANOVA and linear regression of the gene expression with functional hearing measures, and Limma Analysis (Blue). Key genes identified from three different statistical analysis are: CathepsinS (*Ctss*) upregulated in all age groups, Casein kappa (*Csnk*) upregulated in mild and severe presbycusis and down regulated in middle aged mice, Macrophage expressed gene 1 (*Mpeg1* – no primer/ probes were available at the time of the investigation), and C-type calcium dependent sugar recognition protein (*Clecsf12*) which was upregulated in all age groups when compared to young adult mice with good hearing.
Abbreviations: ANOVA, analysis of variance; Limma, linear models for microarray data; RMA, robust multi-chip analysis.

ratio. The error rates were obtained after removing three outliers. The mean error rate (average proportion of misclassified test observations) and standard error for Model 1 and Model 2 were computed (Table 2). The results indicated that gene selection had a lower error rate using the latter model. With Model 2, and a splitting ratio of 2:1, the following 10 genes were selected by RF, in order of importance based on mean decrease of accuracy: *Mpeg1* (macrophage-expressed gene 1), *Ms4a6d* (membrane-spanning 4-domains, subfamily A, member 6D), *Clca1* (chloride channel calcium-activated 1), *Ctss* (cathepsin S), *Bcl2a1a* (B cell leukemia/lymphoma 2-related protein A1a), *Clecsf12* (C-type [calcium dependent, carbohydrate recognition domain] lectin, superfamily member 12), *Cxcl13* (chemokine [C-X-C motif] ligand 13), *Fcgr3* (Fc receptor, IgG, low

affinity III), *Cd68* (CD68 antigen), and *H2-K1* (histocompatibility 2, K1, K region).

An additional search for interesting genes was made using the *P* value and fold-change threshold. A set of 33 genes ($P < 0.0001$ and fold-change >1.55, see Table 3), were

Table 2 Assessing the relevancy of gene expression: random forest cross-validation of the classification for two models

Ratio	Model 1	Model 2
4:1	0.065 ± 0.00274	0.0375 ± 0.0191
2:1	0.107 ± 0.286	0.0308 ± 0.0126

Notes: Model 1: all probe sets only; Model 2: all probe sets, age, gender, and hearing measurements. The mean error rate (average proportion of misclassified test observations) and standard error for different splitting size ratio of the samples are displayed for each model.

Table 3 Limma analysis shows 33 potentially interesting genes from a biological standpoint. The genes were differentially expressed between mice with good hearing and mice with presbycusis, having simultaneously high moderated t statistics (rank used in the table) and high (absolute) log fold-changes

Gene ID	Gene name	Gene function	Fold-change	Moderated t statistics
Upregulated				
1420804_s_at	Clecsf8	C-type (calcium dependent, carbohydrate recognition domain) lectin, superfamily member 8	0.96	13.08
1427076_at	Mpeg1	Macrophage-expressed gene 1	1.28	11.79
1448591_at	Ctss	Cathepsin S	0.91	10.69
1420699_at	Clecsf12	C-type (calcium dependent, carbohydrate recognition domain) lectin, superfamily member 12	1.23	10.17
1419735_at	Csnk	Casein kappa	2.15	9.81
1449164_at	Cd68	CD68 antigen	1.19	9.75
1448620_at	Fcgr3	Fc receptor, IgG, low affinity III	0.97	9.68
1419627_s_at	Clecsf10	C-type (calcium-dependent, carbohydrate recognition domain) lectin, superfamily member 10	0.76	9.45
1419004_s_at	Bcl2a1a	B cell leukemia/lymphoma 2 related protein A1a	0.93	9.18
1435477_s_at	Fcgr2b	Fc receptor, IgG, low affinity IIb	0.71	9.04
1419599_s_at	Ms4a6d	Membrane-spanning 4-domains, subfamily A, member 6D	0.98	8.84
1436905_x_at	Laptm5	Lysosomal-associated protein transmembrane 5	1.17	7.62
1426808_at	Lgals3	lectin, galactose binding, soluble 3	1.52	7.5
1436996_x_at	Lyzs	Lysozyme	0.78	7.2
1450792_at	Tyrobp	TYRO protein tyrosine	0.64	6.55
1415983_at	Lcp1	Lymphocyte cytosolic protein 1	0.68	6.36
1419463_at	Clca1	Chloride channel calcium activated 1	1.22	6.32
1437726_x_at	C1qb	Complement component 1, q subcomponent, beta polypeptide	1.22	6.18
1423547_at	Lyzs	Lysozyme	1.47	6.03
1420394_s_at	Gp49b	Glycoprotein 49 B	0.76	5.78
1435290_x_at	H2-Aa	Histocompatibility 2, class II antigen A, alpha	0.77	5.68
1434366_x_at	C1qb	Complement component 1, q subcomponent, beta polypeptide	0.72	5.61
1422903_at	Ly86	Lymphocyte antigen 86	0.66	5.47
1417851_at	Cxcl13	Chemokine (C-X-C motif) ligand 13	0.98	5.44
1452016_at	Alox5ap	Arachidonate 5-lipoxygenase activating protein	0.84	5.34
1451860_a_at	Trim30	Tripartite motif protein 30	0.65	5.34
1423100_at	Fos	FBJ osteosarcoma oncogene	1.24	5.32
1422124_a_at	Ptprc	Protein tyrosine phosphatase, receptor type, C	1.06	5.14
1427660_x_at	Igk-V28	Immunoglobulin kappa chain variable 28 (V28)	1.53	4.77
1439426_x_at	Lzp-s	P lysozyme structural	1.17	4.68
Downregulated				
1434745_at	Ccnd2	Cyclin D2	−0.65	−4.62
1423523_at	Aass	Aminoadipate-semialdehyde synthase	−0.85	−4.77
1439148_a_at	Pfkl	Phosphofructokinase, liver, B-type	−0.67	−4.83

extracted and ranked. Among them, 30 genes had positive log-fold changes, thus indicating overexpression of these genes with age and hearing loss, and three genes were downregulated (Figure 2). A strong association between these genes (Table 3) and presbycusis is displayed, in order of statistical significance.

Pathway analysis

To determine possible molecular mechanisms underlying presbycusis, pathways analysis was conducted. The first analysis was done with a random forest algorithm (PathwayRF), a pathway-based classification for understanding biological significance in high-throughput gene expression studies. Three outliers were detected. To identify important pathways, 538 pathways from different sources with at least two genes in each pathway were investigated. Table 4 displays the number of genes in each pathway where the out-of-bag (OOB) rate (<5.26%) provides an estimate of classification error with and without outliers. A small OOB error rate is an indication of a pathway being able to classify age-related hearing loss status effectively. The majority of these pathways are related to immune function (B lymphocyte, calorie restriction immune function), being local acute inflammatory responses (classicPathway, lairPathway); cytokine activation leading to
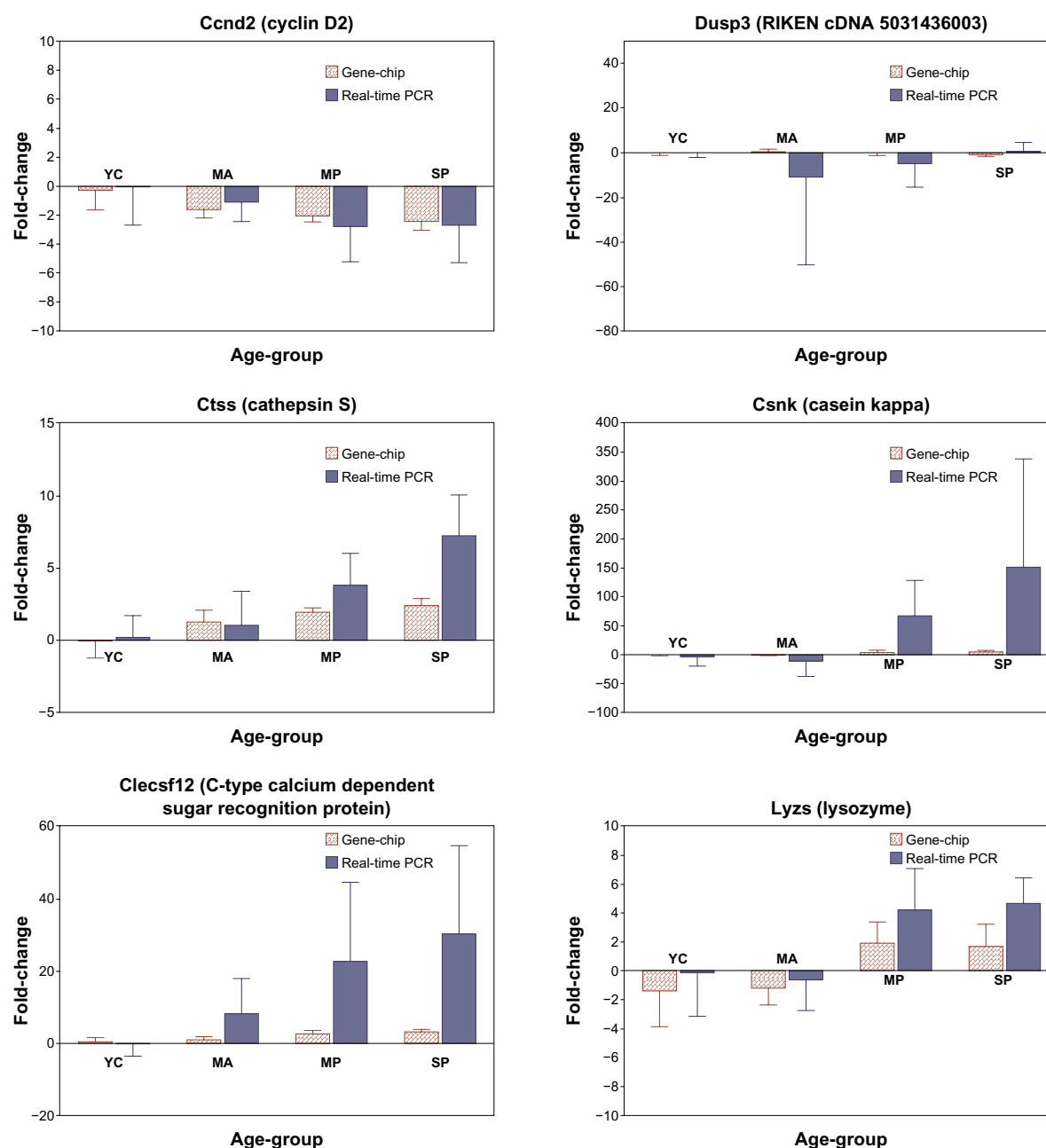
Figure 2 Relative real-time polymerase chain reaction confirmation of GeneChip gene expression RNA was isolated from each mouse cochlea, reverse transcribed to cDNA, and relative real-time polymerase chain reaction was performed on triplicates of each sample. The red bar represents the average mean of GeneChip expression, ie, control eight GeneChips, MA 17 GeneChips, MP nine GeneChips, and SP six GeneChips. The blue color represents the average mean of each mouse cDNA (triplicate) for the relative real-time polymerase chain reaction fold-change.
Note: Error bar represents standard deviation.
Abbreviations: YC, young adult controls with good hearing; MA, middle-aged; MP, mild presbycusis; SP, severe presbycusis; PCR, polymerase chain reaction.

downregulation of cell surface receptors (SA MMP cytokine connection), and epithelial mesenchymal transition (EMT_UP), a complex process that involves changes in gene expression, cytoskeletal reorganization, cell adhesion, extracellular matrix composition, and late-stage tumor progression, and cancer (K-RAS). The next top set of 11 pathways with OOB less than 10% are BADPATHWAY (BioCarta), complement activation classical (gene ontology), compPathway (Biocarta),

etsPathway(BioCarta), shh_lisa (manually curated), RAP-UP, normal MPD (myeloproliferative disorder) BEN, HL60 ATRA BEN, up MEGS, and runx1_targets, cbfb_myh11_downing_aml. The last five pathways were manually curated.

RFs also provide pairwise proximity measures of classes where outliers are defined as cases with small proximity values to most other cases. After removing three such outliers, a multidimensional scaling plot (Figure 3) of proximity

Table 4 Important pathways identified using the random forest method ranked by out-of-bag error rates of $\leq 5.26\%$ without outliers

Pathway database	Pathway name	Description	Genes (n)
BioCarta	B lymphocyte pathway	Humoral immune function	19
	Classic pathway	Local acute inflammatory response	12
	Lair pathway	Local acute inflammatory response	27
PNAS	CR immune function	Humoral immune function	54
SigmaAldrich	SAA MMP cytokine connection	Cytokine activation leading to downregulation of cell surface receptors	20
Jechlinger et al[34]	EMT UP	Epithelial mesenchymal transition	82

Abbreviations: CR, caloric restriction; EMT UP, epithelial to mesenchymal transition upregulation; MMP, matrix metalloproteinase; N, number; PNAS, *Proceedings of the National Academy of Sciences of the United States of America*; SAA, serum amyloid A protein.

matrix shows that the two groups were well separated based on the B lymphocyte pathway (used for illustration) with a 0% OOB rate. When the probes in this pathway were ranked by their t statistic value, four probes were well separated by hearing status, ie, (Ptprc) protein tyrosine phosphatase, receptor type C, (*Icam1*) intercellular adhesion molecule, (*Itgb2*) integrin beta2, (*Fcgr2b*) Fc receptor, IgG, low affinity.

After 15 runs with the same parameters and different seed values, there were seven genes that appeared frequently among the most important genes ranked by RF, ie, *Clec4d*, *Fcgr2b*, *Mpeg1*, *Fcgr3*, *Bcl2a1a*, *Ctss*, and *Htr2b*. They were ranked as importantly as age based on the mean decrease in accuracy. By several measures, the most significant upregulated gene in mice with severe presbycusis is *Clec4d*. This gene was not annotated in any of the gene set categories used for PathwayRF. However, it is annotated with SuperArray and implicated in dendritic antigen presentation and sugar binding with the sigPathway. The next gene, *Fcgr2b* (upregulated with age and hearing loss), is implicated in numerous pathways,

Figure 3 Multidimensional scaling plot of the cochlea of the CBA mouse for B lymphocyte pathways. Coordinate 1 displays the proximities and coordinate 2 displays the distances among samples (N, good hearing group; P, age-related hearing loss group). The two groups were well separated.

including a pathway named cell surface, dendritic-antigen presenting cells, and regulation of immune response. It is included in the B lymphocyte pathway, caloric restriction immune function, normal MPD ben, and KRAS final MIT with PathwayRF. Upregulated *Fcgr3* is also involved in several pathways, including antigen presentation, chemotaxis, and the inflammatory response. It was also found that *Fcgr3* is implicated in SA_MMP_Cytokine connection (PathwayRF), also called cytokine production in sigPathway.

Gene ontology analysis was carried out with the GOstats package in Bioconductor.

Overrepresented annotations in the analyzed list of genes were extracted from the gene ontology database. For the biological process ontology, the gene universe contained 4461 genes mapped to ENTREZ genes. A simple t-test identified 378 differentially expressed genes with P values less than 0.01. The gene ontology analysis includes enhancement of several biological processes, including immune response, chemotaxis, antigen processing and presentation of peptide antigen via the major histocompatibility complex class, and phagocytosis engulfment.

The last pathway analysis tool was applied, for similarity and complementary information. It is an improvement of the gene set enrichment method implemented in the Sig-Pathway Bioconductor package. This tool did not use any gene preselection. It assessed the differential expression of each gene set with respect to hearing loss status between all gene sets and within each gene set. The control probe sets were removed, nonspecific filtering was done (interquartile range < 0.25) and the same three outliers (detected from PathwayRF) were excluded from the analysis. Gene set categories having a minimum of 20 genes and a maximum of 500 genes were searched and the top pathways were found. A strong statistical differential expression of these gene sets was confirmed, with high near-threatened and not evaluated values along with corresponding q values. These data characterize numerous pathways that may play a role in
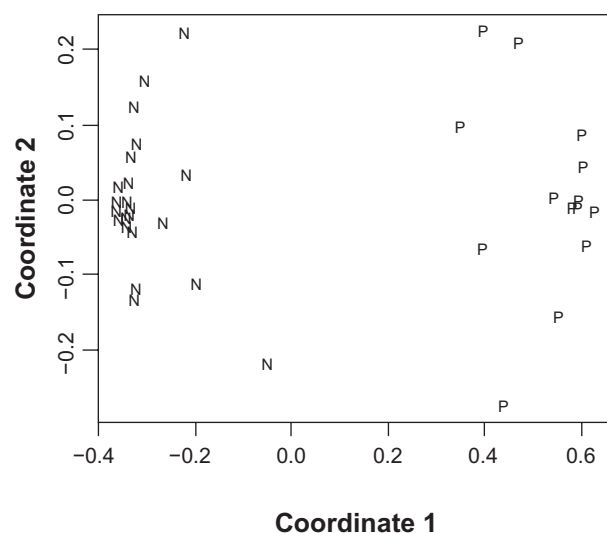
Table 5 Pathway analysis using sigPathway method

Combined rank	Gene set category	Pathway	Set size	Percent changing	NTk stat	NTk rank	NEk stat	NEk rank
1	SuperArray	Dendritic antigen-presenting cell	73	34	10.53	3	5.74	4
2	GO:0009897	External side of plasma membrane	34	18	9.89	5	5.86	2
3	GO:0051707	Response to other organism	112	31	12.06	1	5.62	7
4	GO:0009613	Response to pathogen or parasite	104	32	10.71	2	5.26	10
5	GO:0009986	Cell surface	48	21	9.7	6	5.63	6
6	GO:0019882	Antigen presentation	25	8	10.42	4	5.13	11
7	GO:0006959	Humoral immune response	28	29	7.8	10	5.69	5
8	GO:0005529	Sugar binding	48	42	6.52	20	6.23	1
9	GO:0030333	Antigen processing	21	10	8.73	7	4.51	16
10	GO:0050776	Regulation of immune response	31	19	7.4	14	5.51	9
11	GO:0051240	Positive regulation of physiological processes in organism	24	21	6.28	21	5.81	3
12	GO:0006935	Chemotaxis	37	22	7.7	12	4.94	13
13	GO:0042330	Taxis	37	22	7.7	12	4.94	13
14	GO:0006954	Inflammatory response	44	32	7.49	13	4.94	14
15	GO:0009611	Response to injury	79	39	7.7	11	4.3	19
16	KEGG:04610	Complement and coagulation cascades	27	26	7.28	15	4.76	15
17	GO:0050778	Positive regulation of immune response	22	23	5.95	25	5.6	8
18	GO:0007626	Locomotor behavior	62	32	6.75	17	4.34	18
19	GO:0051239	Regulation of physiological processes in organism	73	38	6.65	19	4.48	17
20	GO:0009605	Response to external stimulus	126	44	8.16	8	3.84	34
21	KEGG:04060	Cytokine–cytokine receptor interaction	83	40	8.03	9	3.93	33
22	SuperArray	Inflammatory cytokine receptors	28	39	5.97	24	4.24	22
23	GO:0001816	Cytokine production	22	23	5.49	34	5	12
24	GO:0001664	G-protein coupled receptor binding	21	19	6.04	23	4.17	24
25	GO:0004896	Hematopoietin/interferon class (D200-domain) cytokine receptor activity	20	25	5.72	30	4.24	21
26	SuperArray	Chemokines and receptors	23	22	5.74	28	4.07	27
27	GO:0004866	Endopeptidase inhibitor activity	56	41	5.61	32	4.17	25
28	KEGG:04650	Natural killer cell-mediated cytotoxicity	49	49	5.72	29	3.83	36
29	SuperArray	Cardiovascular disease	67	43	7.01	16	3.54	60.5
30	GO:0030414	Protease inhibitor activity	57	42	5.2	43	3.83	35
31	GO:0046649	Lymphocyte activation	33	33	5.21	42	3.72	41
32	GO:0045321	Immune cell activation	39	36	5.43	36	3.58	57
33	GO:0001775	Cell activation	39	36	5.43	36	3.58	57
34	GO:0016874	Ligase activity	151	79	−4.84	47	−3.72	47
35	GO:0030246	Carbohydrate binding	85	45	5.71	31	3.43	73.5
36	GO:0001584	Rhodopsin-like receptor activity	36	36	5.5	33	3.43	73.5
37	KEGG:00760	Nicotinate and nicotinamide metabolism	27	89	−3.54	78.5	−3.98	30
38	GO:0048534	Hemopoietic or lymphoid organ development	55	40	5.42	37	3.43	73.5
39	GO:0006461	Protein complex assembly	83	83	−3.65	74	−3.74	39
40	SuperArray	G-protein coupled receptors signaling pathway finder	32	34	5.27	40	3.43	73.5
41	GO:0030695	GTPase regulator activity	93	77	−3.19	97	−4.25	20
42	GO:0016879	Ligase activity, forming carbon-nitrogen bonds	87	75	−3.35	88.5	−3.81	37
43	GO:0019867	Outer membrane	40	88	−3.06	107.5	−3.98	31
44	GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	34	21	5.93	27	3.19	120
45	GO:0004842	Ubiquitin-protein ligase activity	67	75	−3.06	107.5	−3.73	40
46	GO:0005083	Small GTPase regulator activity	57	77	−2.85	123	−4.07	28
47	GO:0007610	Behavior	95	44	5.45	35	3.19	120
48	GO:0016881	Acid–amino acid ligase activity	73	74	−3.04	111	−3.72	47
49	GO:0005813	Centrosome	27	78	−2.59	148	−3.95	32
50	KEGG:04514	Cell adhesion molecules	71	41	6.28	22	2.97	159
51	GO:0005815	Microtubule organizing center	31	74	−2.41	169	−4.03	29
52	GO:0000151	Ubiquitin ligase complex	24	79	−2.48	160	−3.81	38
53	KEGG:04612	Antigen processing and presentation	46	39	5.41	38	2.78	187

(Continued)

Table 5 (*Continued*)

Combined rank	Gene set category	Pathway	Set size	Percent changing	NTk stat	NTk rank	NEk stat	NEk rank
54	GO:0031072	Heat shock protein binding	26	81	−2	223	−4.2	23
55	GO:0043292	Contractile fiber	45	36	5.25	41	2.61	211
56	GO:0030017	Sarcomere	41	32	5.4	39	2.58	222
57	GO:0005764	Lysosome	68	41	6.75	18	2.25	278
58	GO:0000323	Lytic vacuole	68	41	6.75	18	2.25	278
59	GO:0016407	Acetyltransferase activity	27	78	−1.51	309	−4.16	26
60	GO:0005773	Vacuole	82	46	5.95	26	1.59	391
61	GO:0048475	Coated membrane	20	75	−1.08	407	−3.72	47

Notes: The gene sets were ranked according to the average of NTk and NEk statistics; percent changing represents the percentage of genes in the corresponding pathway changing expression with presbycusis.

Abbreviations: GTPase, guanosine triphosphatase; Stat, statistic.

presbycusis, ie, they involve genes, the expression of which is significantly different between the two groups (Table 5). The top pathway involved antigen-presenting dendritic cells, which are believed to be involved in autoimmune disease and neuroinflammatory disorders. In each gene set, on average, 40% of the genes change expression with presbycusis and the majority of the gene sets are upregulated (indicated by a positive near-threatened and not evaluated statistic). In the SigPathway method, "percent up" indicates the percentage of genes in the set that were more greatly changed by presbycusis in the nonhearing group. The analysis indicated a major disturbance (downregulation) of heat shock protein binding (81% percent up), whereas GTPase regulator activity (77% percent up) increased with presbycusis. The majority of gene sets differentially expressed were related to immune function, such as traditional antigen-presenting cells, proteinases involved in inflammatory disease, and immunoinhibitory receptors. Each gene found differentially expressed with Limma analysis was listed in at least one of the 32 top gene sets.

Confirmation of array probe set by relative real-time PCR

We compared the microarray expression levels of a subset of genes with results from an independent measure of gene expression using relative real-time polymerase chain reaction (PCR) derived from the same RNA, but independent cDNA from those used for the microarrays (Figure 2). The four categories of mice based on age and hearing ability were considered. When comparing aged mice with severe presbycusis and the young controls, the genes found to be important from the statistical analysis were significantly upregulated ($P < 0.001$), with fold-changes at least 2.0 (above) or −2 (below) relative to the young control group. As shown in Figure 2, relative real-time PCR analysis of the reverse transcribed RNA of six probe sets showed that all age groups

in each probe set conformed to the direction of the GeneChip gene expression. In six probe sets used for real-time PCR, basically three patterns of fold-change was observed, either all age group gene expression increased or decreased or there was a mixture of increase and decrease in the same probe set. *Cyclin D2* was downregulated in all age groups. *Ctss* and *Clecsf12* were upregulated in each age group. RIKEN cDNA 5031436003 (*Dusp3*) was downregulated in middle-aged mice with good hearing and old-aged mice with mild presbycusis, and upregulated in severe presbycusis. *Csnk* and *Lysozyme* were downregulated in middle-aged mice with good hearing, and were upregulated in old-aged mice with mild presbycusis or severe presbycusis when compared with young adults with good hearing. Notably, the genes included encoding proteins that are usually implicated in the extracellular space, hydrolase and antimicrobial activity, polysaccharide (zymosan) receptor-binding activity, and cysteine-type endopeptidase activity.

Finally, genes that are upregulated in the aged mice with mild presbycusis and severe presbycusis groups (*Csnk* and *Lysozyme*) are associated with peptidase activity molecules connected with cellular reorganization. Overall, these data suggest increased activation of a wide range of cellular cochlear components in older age groups when compared with young adult controls with normal hearing. On the other hand, *cyclin D2*, a gene associated with regulation of the cell cycle and cytokinesis, was downregulated in middle-aged mice with good hearing, as well as aged mice with mild presbycusis or severe presbycusis when compared with young controls.

A dendogram was constructed to help integrate and summarize the findings. It is noted that the set of genes in the dendogram overlaps with the 10 genes from Model 2 (splitting ratio 2:1) and the 89 genes from the Limma analysis (Figure 4). The dendogram on the side was obtained from an agglomerative complete linkage algorithm using Euclidian distance (Figure 5).
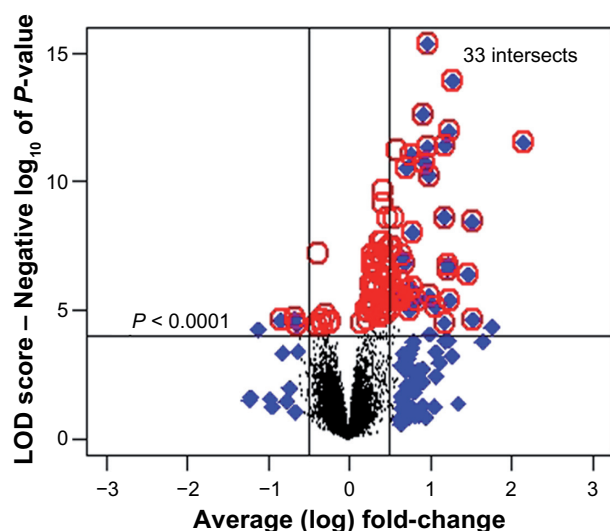
Figure 4 Volcano plot with moderated *t* statistics: 33 intersecting genes (blue diamonds within red circles) show potentially interesting genes using the following criteria: adjusted *P* < 0.0001 and relatively high absolute average (log) fold-change > 0.5. Genes represented by the black data points do not possess these criteria.

Promoter sequence analysis

A promoter module is defined as a functional unit consisting of two or more transcription factor-binding sites conserved in both order and distance. Promoter analysis of the *Ctss*, *Csnk*, *Mpeg1*, and *Clecsf12* mouse genes was performed using Genomatix ElDorado™, as described by Chen et al.[29] Interestingly, all four top genes in the present study which alter their expression with age-related hearing loss have the same GXP promoter. Promoter GXP, which consists of about 600–800 base pairs, is located at GXP_161004 in the *Ctss* gene, promoter GXP 152047 in *Csnk*, promoter GXP_304631 in *Mpeg1*, and promoter GXP_438841 in *Clecsf12*. Furthermore, the DiAlign transcription factor task of the GEMS Launcher (MatInspector)[30] was used to check for conserved transcription binding factors from six species. In the present investigation, this analysis predicted transcription factor module MYOD_VTBP_01 in the *Ctss* gene and transcription factor module MYOD_SRFF_01 in the *Csnk* gene. This promoter is conserved in six species and codes for 1–6 relevant transcripts in the *Ctss*, *Csnk*, *Mpeg1*, and *Clecsf12* genes.

Discussion
Previous analyses

We used DNA microarrays and quantitative PCR to examine gene expression profiles in aging CBA/CaJ mice raised under normal conditions. Although a good model of the typical time course of progression of presbycusis, no animal model
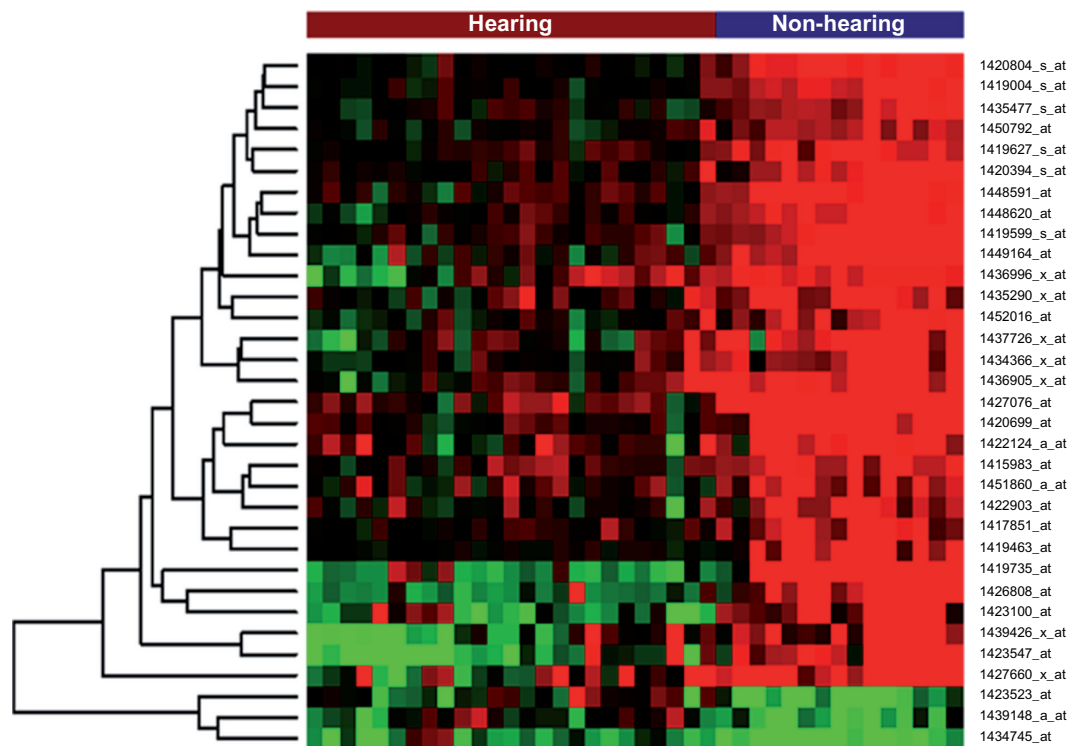

Figure 5 Heatmap of robust multichip average normalized expression of 33 differentially intersecting expressed genes, with statistically significant adjusted *P* values and high absolute fold-changes, overexpressed (red), and underexpressed (green) with age and hearing loss. The expression profiles were for samples from 25 hearing mice and 15 mice with age-related hearing loss. The dendrograms at the side were constructed from a hierarchical clustering with an agglomerative complete linkage algorithm, using a Euclidean distance measure. Genes with black cells for a sample did not show any change.

can capture all aspects of cochlear presbycusis and aging of the central auditory system. In addition, the novel findings reported here for new genes that may be involved in cochlear presbycusis can serve as a useful guide for geneticists in search of human genes that may predispose to, accelerate, or protect aged individuals against severe age-related hearing loss.

A previous report analyzed the same microarray data using a novel multivariate search for differentially expressed genes where two-sample hypotheses were formulated to test the joint distribution of the subvector of expression signals.[14] The outcome of this analysis resulted in identification of 10 probe sets (as shown in the Venn diagram of Figure 1). For the sets exhibiting major expression changes with age, six probe sets were functionally related to the immune system, three to post-translational modification, and one to mammary gland function. In another simpler statistical approach,[10] the normalized data were subjected to a three-criteria selection protocol: one-way analysis of variance for probe sets with $P < 0.0001$ and better; linear regression of functional hearing data (ABR, DPOAE) versus gene expression (GeneChip data); and a high fold-change cutoff of above 2 or below −2. When the data, consisting of 22,262 probe sets, were analyzed according to this protocol, it was found that 34 genes were upregulated and 26 were downregulated with age and hearing loss, as shown in Venn diagram (Figure 1).

Present analysis
In the present study, global analysis of covariance indicated that the overall gene-expression profile is associated with the hearing status using two groups (hearing versus nonhearing). Thus, a binary grouping was used to detect global differences in gene expression adjusted for subject variables. A two-step approach (PLS and RF) determined that microarray data were not redundant in predicting hearing status. Combining these two data analysis tools was quite effective. The first component extracted from PLS was ranked next to age and was better than subject characteristic predictors. After this finding, RF alone was used to obtain the specific probe sets that performed equally well or better than age and other covariates in the context of prediction and classification. The approach yielded good prediction accuracy based on the mean error rate. It was successful in separating age and probe sets that best differentiate genes for which expression changed with age-related hearing loss. Variable importance measures obtained from RF classification provided a ranked set of predictors

(probes and covariates). Neurobiologically speaking, it was interesting that none of the DPOAE measures made the top list. Given that DPOAEs measure outer hair cell function, this suggests that the gene expression changes with age and hearing loss are probably occurring in cochlear sensory cells other than the outer hair cells, such as inner hair cells, spiral ganglion neurons, or cells of the stria vascularis or spiral ligament. This is a possibility, because the mouse cochlear tissue samples of the present investigation included cells from all of these areas.

The gene-based analysis moderated t-test identified 89 differentially expressed genes (adjusted for $P < 0.01$), of which four are common to all statistical methods used for these data. These four genes are involved in tumor formation, inflammation processes, autoimmune disease, and cytokine production (immune response). *Ctss* is a lysosomal cysteine endoprotease which fosters antigen presentation. It plays an important role in the regulation of major histocompatibility complex II expression, a biological process which apparently is upregulated during presbycusis (gene ontology analysis). This shows evidence of inflammatory immune pathway activation that occurs in the cochlea with age and hearing loss. *Csnk* was highly upregulated. It is involved in mammary gland functioning and may not be directly involved in hearing loss. This change might be explained as reflecting some degenerative process in the inner ear in older female mice with lower levels of estrogen and progesterone. *Mpeg1* is found at high levels in malignant progression of tumors. *Clecsf12* is a C-type lectin receptor. It is selectively expressed in dendritic cells of the immune system (a set of genes identified by the sig-Pathway method). However, it should also be considered that the immune-related pathways implicated in the present study may reflect only responses to cell death in old mice, rather than being the fundamental cause.

Real-time PCR analyses confirmed the gene expression levels (Figure 2) found using the microarray screen. Analysis of expression of individual genes and gene sets, and of pathways that were upregulated or downregulated in the old mild and severe presbycusis subject groups led to several new conclusions. For example, the hearing loss accompanying aging is likely to be related to B lymphocyte increases in IgV8 and a decrease in cyclin D2 associated with the cell cycle and cytokinesis.

Our findings point to coordinated sets of genes and pathways that are linked to presbycusis. The RF analysis identified the B lymphocyte pathway as being important in age-related hearing loss.

Several previous studies of presbycusis are consistent with the present investigation. Hearing loss has been linked to a condition known as autoimmune sensorineural hearing loss. The organ of Corti is the crucial tiny structure inside the inner ear containing the ultrasensitive hair cells which respond to vibrations by creating movement and sending neural signals to the brain, eventually interpreted as sounds such as speech. With age, the immune system can damage the organ of Corti, including the hair cells, and diminishing or destroying auditory function. This immune-related damage has been implicated in other neurodegenerative diseases of aging, and may be an inevitable part of the mammalian aging process itself. For example, Alzheimer's disease, atherosclerosis, and type II diabetes, have been linked to age-related autoimmune diseases.[31] Previous studies have also demonstrated that the immune system attacks the inner ear. In humans, autoimmune damage to the inner ear is suspected to be a cause of rapidly progressing hearing loss.[32] The gene set analysis in this study confirms the important role of immune system function in the pathogenesis of presbycusis.

Moreover, the loss of hair cells is a common etiology for hearing loss in older people. Damage to these hair cells cannot be repaired, and loss of these sensory cells causes permanent hearing loss or deafness in mammals because, unlike birds and lower vertebrates, mammalian hair cells are generated only during embryogenesis.

The present study also found that certain aspects of presbycusis are related to the epithelial-to-mesenchymal transition pathway. The EMT_UP pathway (one of the top six pathways with PathwayRF) is associated with inner ear hair cell development. When cells undergo an epithelial-to-mesenchymal transition, they differentiate in a mesenchymal form and supporting cells can increase in number via mitotic proliferation. Some of the cells will begin to express hair cell-specific markers, but the cells will not develop definitive hair cell morphology. If these cells are forced to undergo a mesenchymal-to-epithelial transition, spheres that include both hair cells (expression of a hair cell marker, myosin VIIa) and supporting cells will be produced.[33]

The gene *Bcl2a1a* was not in any of the gene set categories used. However, it is upregulated and related to apoptosis with the biological process ontology. It is evident that apoptosis is involved in deafness, and especially in presbycusis of cochlear origin.[13] A supervised learning approach gave an error rate estimate of the classification after cross-validation and determined the usefulness of gene expression. Additionally in order to extract the most informative genes and covariates explicitly, the RF method alone was applied.

On the basis of the results, real-time PCR assays for a range of cDNA in organ of Corti-expressed genes were used to validate data from the microarrays (Figure 2). We found that presbycusis specifically downregulates *cyclin D2* in middle-aged mice with good hearing when compared with young control mice with good hearing. Furthermore, presbycusis upregulated a wide array of genes (*Ctss*, *Clecsf12*, *Csnk*, and *Lysozyme*) in mice with mild and severe presbycusis. A major finding is that a key pathway not previously implicated in age-related hearing loss, ie, the B lymphocyte pathway, may play an important role in presbycusis.

Although promoter GXP was not the focus of the present investigation, a preliminary promoter sequence analysis was initiated to spur future investigations. The prediction of the same GXP promoter module in all four of the top genes identified in the present investigation (*Ctss*, *Csnk*, *Mpeg1*, and *Clecsf12*) was discovered, and analyzed using Genomatix (ElDorado). In addition, further study of the predicted transcription factor module of MYOD_VTBP_01 in the *Ctss* gene and transcription factor module MYOD_SRFF_01 in the *Csnk* gene would be an interesting approach to help identify potential regulatory pathways linked to the progression of presbycusis.

Conclusion

In this gene expression study, we determined the predictive value of gene expression in addition to gender and hearing measurements. Using a novel classification approach, the most important probes that predict hearing loss besides age were identified. The CBA mouse model utilized here has proven to be very useful for examining the neural and molecular bases of presbycusis, given that this mouse strain exhibits a slow, progressive hearing loss relatively late in life, which is similar to most humans, when one corrects for the different absolute lifespan of mice and humans. The results showed that gene expression alterations were linked to age and hearing loss in the mouse cochlea. Limma analysis identified 89 genes, of which 33 are potentially interesting genes that changed expression in conjunction with presbycusis, based on large moderated *t* statistics and high fold-changes (Figure 4). It included one gene discovered in a previous single gene analysis and five genes (from a set of 10) from a previous study of a gene set from the same high-throughput data set. However, ranking the genes by a univariate test statistic ignores the dependence among genes and gene-to-gene interactions. This report demonstrates the value of

interpreting the gene expression changes within the context of molecular signaling and regulatory pathways to explain the biological mechanisms underlying presbycusis. To ensure that different possible important pathways were identified, three different tools were used to maximize information on the biological relevance. The findings indicated that multiple pathways might have a significant role in presbycusis, including the immune system, dendritic antigen-presenting cells, carbohydrate metabolism, epithelial-to-mesenchymal transition, complement and coagulation cascades, chemokines, and G-protein coupled receptor binding. Although the most interesting genes and the top pathways identified had known biological significance, there were a few top genes not included in the existing annotation databases. Promoter analysis by Genomatix revealed an interesting promoter, GXP, which is present in all four top genes reported here as significantly changing their expression in age-related hearing loss. Our findings allow uncovering of novel cross-linking pathways, or genes that were not found by prior studies, including those with previously unknown branching in the gene networks.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Frisina DR, Frisina RD, Snell KB, Burkard R, Walton JP, Ison JR. Auditory temporal processing during aging. In: Hof PR, Mobbs CV, editors. *Functional Neurobiology of Aging*. San Diego, CA: academic press; 2001.
2. Ohlemiller KK, Frisina RD. Clinical characterization of age-related hearing loss and its neural and molecular bases. In: Schacht J, Popper A, Fay RR, editors. *Auditory Trauma, Protection and Treatment*. New York, NY: Springer-Verlag; 2008.
3. Pollak GD, Park TJ. The effects of GABAergic inhibition on monaural response properties of neurons in the mustache bat's inferior colliculus. *Hear Res*. 1993;65:99–117.
4. Caspary DM, Raza A, Armour BAL, Pippin J, Arneric SP. Immunocytochemical and neurochemical evidence for age-related loss of GABA in the inferior colliculus: implications for neural presbycusis. *J Neurosci*. 1990;10:2363–2372.
5. Caspary DM, Milbrandt JC, Helfert RH. Central auditory aging: GABA changes in the inferior colliculus. *Exp Gerontol*. 1995;30:349–360.
6. Caspary DM, Holder TM, Hughes LF, Milbrandt JC, McKernan RM, Naritoku DK. Age-related changes in GABA(A) receptor subunit composition and function in rat auditory system. *Neuroscience*. 1999;93: 307–312.
7. Caspary DM, Schatteman TA, Hughes LF. Age-related changes in the inhibitory response properties of dorsal cochlear nucleus output neurons: role of inhibitory inputs. *J Neurosci*. 2005;47:10952–10959.
8. Caspary DM, Ling L, Turner JG, Hughes LF. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J Exp Biol*. 2008;211(Pt 11):1781–1791.
9. Oh S, Kim C, Song J. Gene expression and plasticity in the rat auditory cortex after bilateral cochlear ablation. *Acta Otolaryngol*. 2007;127:341–350.
10. D'Souza M, Zhu X, Frisina RD. Novel approach for selecting genes from RMA normalized microarray data using functional hearing tests in aging mice. *J Neurosci Methods*. 2008;171:279–287.
11. Tadros S, D'Souza M, Zettel M, Zhu X, Lynch-Erhardt M, Frisina RD. Serotonin 2B receptor: upregulated with age and hearing loss in mouse auditory system. *Neurobiol Aging*. 2007;28:1112–1123.
12. Tadros S, D'Souza M, Zettel M, Zhu X, Lynch-Erhardt M, Frisina RD. Glutamate-related gene expression changes with age in the mouse auditory midbrain. *Brain Res*. 2007;1127:1–9.
13. Tadros S, D'Souza M, Zhu X, Frisina RD. Apoptosis-related genes change their expression with age and hearing loss in the mouse cochlea. *Apoptosis*. 2008;13:1303–1321.
14. Xiao Y, Frisina RD, Gordon A, Klebanov L, Yakovlev A. Multivariate search for differentially expressed gene combinations. *BMC Bioinformatics*. 2004;5:164–184.
15. Ohlemiller KK, Dahl AR, Gagnon PM. Divergent aging characteristics in CBA/J and CBA/CaJ mouse cochleae. *J Assoc Res Otolaryngol*. 2010;11:605–623.
16. Boulesteix A. PLS dimension reduction for classification of microarray data. Sonderforschungsbereich 386, Paper 392. 2004. Available from: http://epub.ub.uni-muenchen.de/. Accessed June 29, 2011.
17. Boulesteix A, Strimmer K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinformatics*. 2007;8:32–44.
18. Boulesteix A. WilcoxCV: an R package for fast variable selection in cross-validation. *Bioinformatics*. 2007;23:1702–1704.
19. Breiman L. Manual on setting up, using, and understanding Random Forests V4.0. 2003. Available from: ftp://ftp.stat.berkeley.edu/pub/users/breiman/. Accessed June 29, 2011.
20. Christensen N, D'Souza M, Zhu X, Frisina RD. Age-related hearing loss: aquaporin 4 gene expression changes in the mouse cochlea and auditory midbrain. *Brain Res*. 2009;1253:27–34.
21. Irizarry RA, Bolstad BM, Collin F. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31:e15.
22. Livak, KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*. 2001;25:402–408.
23. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York, NY: Springer; 2005.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995;57:289–300.
25. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5:R80.
26. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–29.
27. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering statistically significant pathways in expression profiling studies. *Proc Nat Acad Sci U S A*. 2005;102:13544–13549.
28. Subramanianan A, Tamayoa P, Moothaa VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profile. *Proc Nat Acad Sci U S A*. 2005;102: 15545–15550.
29. Chen Y, Lin G, Huo JS, et al. Computational and functional analysis of growth hormone (GH)-regulated genes identifies the transcriptional repressor B-cell lymphoma 6 (Bc16) as a participant in GH-regulated transcription. *Endocrinology*. 2009;150:3645–3654.

30. Cartharius K, Frech K, Grote K, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*. 2005;21:2933–2942.

31. Larbi A, Fiilop T, Pawelec G. Immune receptor signaling, aging and autoimmunity. *Adv Exp Med Biol*. 2008;640:312–324.

32. McCabe BF. Autoimmune sensorineural hearing loss. *Ann Otol Rhinol Laryngol*. 1979;88:585–589.

33. Hu Z, Corwin JT. Inner ear hair cells produced in vitro by a mesenchymal-to-epithelial transition. *Proc Nat Acad Sci U S A*. 2007;104: 16675–16680.

34. Jechlinger M, Grunert S, Beug, H. Mechanisms in epithelial plasticity and metastasis: insights from 3D cultures and expression profiling. *J Mammary Gland Biol Neoplasia*. 2002;7:415–432.