SCIENTIFIC DATA

OPEN

Data Descriptor: Transcriptomes of cochlear inner and outer hair cells from adult mice

Yi Li[1,2], Huizhan Liu[2], Kimberlee P. Giffen[2], Lei Chen[2,3], Kirk W. Beisel[2] & David Z. Z. He[2]

Inner hair cells (IHCs) and outer hair cells (OHCs) are the two anatomically and functionally distinct types of mechanosensitive receptor cells in the mammalian cochlea. The molecular mechanisms defining their morphological and functional specializations are largely unclear. As a first step to uncover the underlying mechanisms, we examined the transcriptomes of IHCs and OHCs isolated from adult CBA/J mouse cochleae. One thousand IHCs and OHCs were separately collected using the suction pipette technique. RNA sequencing of IHCs and OHCs was performed and their transcriptomes were analyzed. The results were validated by comparing some IHC and OHC preferentially expressed genes between present study and published microarray-based data as well as by real-time qPCR. Antibody-based immunocytochemistry was used to validate preferential expression of SLC7A14 and DNM3 in IHCs and OHCs. These data are expected to serve as a highly valuable resource for unraveling the molecular mechanisms underlying different biological properties of IHCs and OHCs as well as to provide a road map for future characterization of genes expressed in IHCs and OHCs.

Design Type(s)	replicate design ● cell type comparison design ● transcription profiling by high throughput sequencing design
Measurement Type(s)	transcription profiling assay
Technology Type(s)	RNA sequencing
Factor Type(s)	cell type
Sample Characteristic(s)	Mus musculus ● inner hair cell ● outer hair cell

[1]Department of Otorhinolaryngology, Beijing Tongren Hospital, Beijing Capital Medical University, Beijing 100730, China. [2]Department of Biomedical Sciences, Creighton University School of Medicine, Omaha, Nebraska 68170, USA. [3]Chongqing Academy of Animal Science, Chongqing 402460, China. Correspondence and requests for materials should be addressed to Y.L. (email: alinaliyi@163.com) or to D.Z.H. (email: hed@creighton.edu)

Background & Summary

Hair cells are the sensory receptors of both the auditory system and the vestibular system in the ears of all vertebrates. Hair cells transduce mechanical stimuli, i.e., movement in their environment, into electrical activity[1,2]. There are two types of hair cells in the mammalian cochlea, inner hair cells (IHCs) and outer hair cells (OHCs). These two types of hair cells are anatomically and functionally distinct[3]. Although much is known about how IHCs and OHCs function in hearing, we have limited knowledge of molecular mechanisms, i.e., gene expression and regulation, that underlie their distinct morphological and functional specializations.

While all cells in multicellular organisms have nearly identical genome, the genes that are transcribed are different for each cell type. Diverse patterns of gene expression and post-transcriptional regulation of gene expression by miRNA underlie phenotypic variances of different cell types. Genome-wide characterization of cell-specific transcriptomes is central to understanding the biological property of a cell or a population of cells. High-throughput mRNA sequencing (RNA-seq) allows simultaneous transcript discovery and abundance estimation with a wide dynamic range and lower false-negative and false-positive discovery rates[4,5]. Direct sequencing of RNA libraries also provides the opportunity to explore alternative splicing, a key mechanism that contributes to transcriptome diversity[6–8]. Transcriptome analysis has emerged as a powerful tool in revealing the genetic and molecular profile of a cell or a population of cells.

In a previous study, we used the microarray technique to examine transcriptional profiling of purified IHCs and OHCs from adult mice[9]. Although microarray is a powerful technique, it has limitations in dynamic range and identification of multiple splice variants of the genes. Furthermore, it relies on prerequisite sequence information, which precludes analysis of unannotated genes[10]. Because of this, 22 to 24% of the transcripts detected in our previous microarray study were uncharacterized or unannotated transcripts or genes[9]. Hair cell-specific transcriptomes have been analyzed using RNA-seq in several recent studies[11–14]. However, these studies analyzed transcriptomes of cochlear and vestibular hair cells only from embryonic and neonatal mice. Furthermore, those studies did not distinguish between IHCs and OHCs.

Here, we describe transcriptome-wide profiling of IHCs and OHCs obtained from one-month-old CBA/J mouse cochleae to provide a comprehensive view of the gene expression in IHCs and OHCs. Unlike some other mouse strains (such as C57/B6) which carry mutations that can cause early onset of age-related hearing loss, CBA/J mice do not exhibit age-related hearing loss until 18 months of age. We took advantage of the established pulled glass pipette technique[9,15] and distinct morphology of the two types of hair cells to separately collect 1,000 isolated IHCs and OHCs. Two biological replicates of IHCs and three replicates of OHCs, each containing 1,000 hair cells, were prepared for RNA-seq. An overview of the study design is depicted in Fig. 1a. Transcriptomes of adult IHCs and OHCs from microarray technique[9], as well as neonatal hair cells from RNA-seq[11–14] were presented along with transcriptomes from the current study. We validated our results by comparing some IHC and OHC preferentially expressed genes between the present study and previous studies[9,11–14] as well as by real-time quantitative PCR (RT qPCR). In addition, we used antibody-based immunostaining to show the preferential expression of SLC7A14 and DNM3, whose function in hair cells has not been characterized. While SLC7A14 showed strong staining in the soma of IHCs, DNM3 was detected in the stereocilia bundle of only OHCs. These two genes/proteins can be used as specific markers for adult IHCs and OHCs. Finally, we examined the expression of deafness-related genes in hair cells. Mutations or deficiencies affecting approximately 150 genes have been linked to inherited syndromic or non-syndromic hearing loss[16]. We analyzed the expression of 143 known deafness genes, excluding X-chromosome-linked genes, and showed 128 genes are expressed in hair cells.

Our dataset is expected to serve as a highly valuable resource for unraveling the molecular mechanisms underlying different biological properties of IHCs and OHCs. The dataset will also provide a road map for future characterization of genes expressed in these two types of hair cells and for assisting the auditory research community in exploring the functions of deafness-related genes.

Methods

Hair cell isolation and collection

CBA/J mice aged between 28 and 35 days old were used for the study. The basilar membrane together with the organ of Corti was isolated as described before[15]. The sensory epithelium was transferred to an enzymatic digestion medium containing 1 ml L-15 and 1 mg Collagenase IV (Sigma) in a small Petri dish. After 5 min for incubation at room temperature (20 ± 2 °C), the tissue was transferred to a small plastic chamber (0.8 ml in volume) containing enzyme-free Leibovitz's L-15 medium (7.35 pH, 300 mOsm). Hair cells were separated after gentle trituration of the basilar membrane with a 200 μL Eppendorf pipette tip. The chamber containing the hair cells was then mounted onto the stage of an inverted Olympus IX71 microscope equipped with a video camera. The chamber (with inlet and outlet) was perfused with fresh L-15 medium to wash out debris for 5 min. IHCs and OHCs in most cases retained their distinct morphological feature after isolation. Some representative images of solitary IHCs and OHCs are presented in Fig. 1b.

To collect solitary hair cells, two pulled glass pipettes with a diameter of ~30 μm were used to pick up and transfer IHCs and OHCs. Each pipette was designated for one cell type to prevent cell type
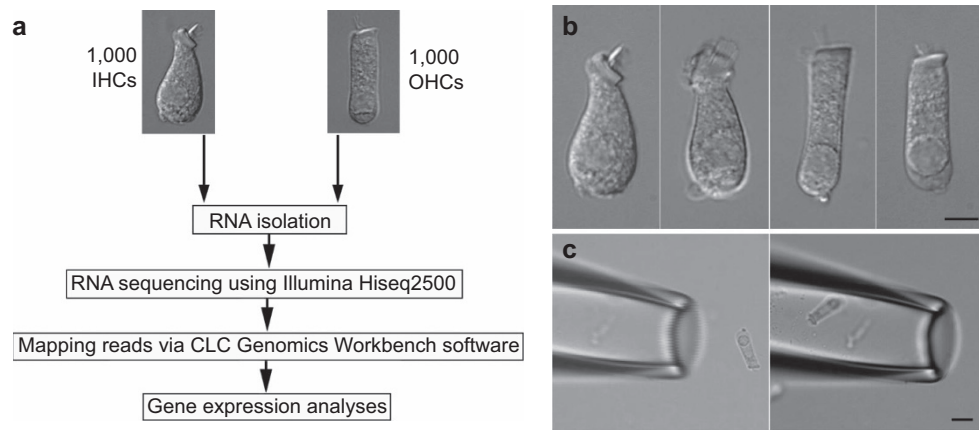
Figure 1. **Study design workflow for RNA-seq and suction pipette technique for collecting isolated hair cells.** (**a**) Workflow of experimental design for RNA-seq transcriptomic analysis of IHCs and OHCs isolated from adult mouse cochleae. (**b**) Representative images (from left to right) of two isolated IHCs and two OHCs from adult mice. (**c**) A pick-up pipette before and after an isolated OHC was drawn into the pipette. This technique was used to individually collect isolated hair cells. Bars: 5 (**b**) and 10 (**c**) μm.

contamination in the pipette. The pickup pipette was fabricated from 1.5 mm thin-wall glass tubing pulled by a two-stage electrode puller. The pipettes were mounted in two separate electrode holders mounted on two Narashigi micromanipulators (Narashigi, Japan). The suction port of the pipette holder, held by the micromanipulator, was connected to a micrometer-driven syringe to provide positive or negative pressure to draw in or expel the cells. An image of an OHC before being drawn into a pickup pipette is shown in Fig. 1c. A video showing a mouse OHC being drawing into a pickup pipette is provided (Data Citation 1). IHCs and OHCs were identified based on their morphology under direct visual observation and solitary hair cells that were not attached to any other cell types were collected. Any hair cells with ambiguous morphology were excluded. Hair cells were transferred to a microcentrifuge tube containing 50 μl RNA*later* (Thermo Fisher Scientific, Waltham, MA) after ~10 cells were collected in the pipette. Cells were expelled from the pipette by applying positive pressure. This step was repeated until approximately 50 to 80 IHCs and 100 to 150 OHCs were collected from each mouse. Thirty mice were used for the collection of two biological replicates of IHCs and three replicates of OHCs.

RNA isolation, amplification
Approximately 1,000 cells suspended in 100 μL RNA*Later* from each biological replicate were used to extract total RNA, including small RNAs (>~18 nucleotides), using the Qiagen miRNeasy mini plus Kit (Qiagen Sciences Inc, Germantown, MD). DNA contamination was eliminated by on-column DNase digestion. The quality and quantity of RNA after purification was examined using an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA) and compared to examples of pure RNA results found in the Agilent 2100 Bioanalyzer 2100 Expert User's Guide. Total RNA from each sample was approximately 8 to 10 ng/μl (with ~3–4 μl total for each sample). These samples were reverse transcribed into cDNA and amplified using the SMART-Seq V4 Ultra Low Input RNA kit (Clontech Laboratories, Inc., Mountain View, CA).

RNA-sequencing and bioinformatic analyses
Genome-wide transcriptome libraries were produced from biological replicates of IHCs and OHCs. SMART-Seq V4 Ultra Low Input RNA kit (Clontech) was used to generate cDNA in combination with the Nextera Library preparation kit (Illumina, Inc., San Diego, CA). To ensure the inserts were the appropriate size and to determine concentration prior to sequencing, a Bioanalyzer 2100 and a Qubit fluorometer (Invitrogen) were used to assess library size and concentration. Transcriptome libraries were sequenced using the HiSeq 2500 Sequencing System (Illumina). Libraries were multiplexed and three samples per lane were sequenced as 100-bp paired-end reads. This generated approximately 100 million reads per sample. The files from the multiplexed RNA-seq samples were demulitplexed and fastq files representing each library and quality control data were generated.

Bioinformatics analyses
CLC Genomics Workbench software (CLC bio, Waltham, MA, USA) was used to map the reads to the mouse genome (mm10, build name GRCm38) and generate gene expression values in the normalized form of reads per kilobase of transcript per million mapped reads (RPKM) values. Reads were mapped to exonic, intronic, and intergenic sections of the genome. Gene expression estimates were derived from the

mapped reads using HTSeq count[17]. Ingenuity IPA program (www.ingenuity.com) and DAVID[18] were used for functional annotation. Entrez Gene, HGNC, OMIM, and Ensembl database were used for verification, reference, and analyses.

Code availability

No custom code was used in any of these analyses.

Real-time qPCR

We validated the expression of 26 genes using RT qPCR. RT qPCR experiments were run on an Applied Biosystems 7500 Fast Real-Time PCR system. Ten microliters of Powerup SYBR Green Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) was used in each 20 microliter reaction. Primer concentrations were 450 nM. The original cDNA samples were diluted twenty-fold with two microliters for every reaction. The fast thermal cycling mode of the Applied Biosystems 7500 instrument was used. We calculated ΔCt values ($\Delta Ct = Ct^{(GOI)} - Ct^{AVG\ HKG}$) of each gene (gene of interest or GOI) after normalizing to Ct value of a house-keeping gene (HKG). For comparing differential expression of a gene between IHCs and OHCs, we calculated $\Delta\Delta Ct$, where $\Delta\Delta Ct = \Delta Ct\ (IHCs) - \Delta Ct\ (OHCs)$[19]. Thus, a positive value would suggest that this gene has a higher expression value in IHCs than OHCs, whereas a negative value suggesting higher expression in OHCs than in IHCs.

The sequences of the oligonucleotide primers were designed using A plasmid Editor (ApE) software (http://biologylabs.utah.edu/jorgensen/wayned/ape/) and BLAST searches (http://blast.ncbinlm.nih.gov/Blast.cgi.) to find unique and appropriate sequences with melting temperatures above 60 °C that had predicted low rates of homodimerization. Oligonucleotide primers were acquired from Integrated DNA Technologies (Coralville, Iowa). The sequences of oligonucleotide primers are shown in Table 1.

GENE	FOWARD PRIMER	REVERSE PRIMER
Bcl2	ACGTGGACCTCATGGAGTG	TGTGTATAGCAATCCCAGGCA
Bcl6	GTG TCC TGG GGT TAC AGG TG	CCT GTC CTG CCT ACC CAT AG
Dnajc5b	ATTTTTGTTGCTGCCTTTGC	AGGCTGGAGAACAACTGGAA
Ednrb	AAAGCCAACGATCACGGATA	CCTTTCTGCTAGCATGGTTTTT
Fzd4	TGCCAGAACCTCGGCTACA	ATGAGCGGCGTGAAAGTTGT
Hdac3	GCCAAGACCGTGGCGTATT	GTCCAGCTCCATAGTGGAAGT
Otof	GGC GCT TCA TIT ATC CTT TCG AC	GAC GAG GTG CCG GAT TGC CTT TAG C
Prkd1	GGGGGCATCTCGTTCCATC	GTGCCGAAAAAGCAGGATCTT
Slc17a8	GGAGACAGAACTCAACCACGA	TTCGGGCCTGGTAGGATAATG
Slc1a3	GCACCAAGTGTTGGAAACTG	TTCAAATGTAGGCTAAAACCGATA
Slc7a14	CACCCTGGTCTCTGTCTGTG	CTGGAAAATTCCTCCCCTTC
Tbx2	CCGATGACTGCCGCTATAAGT	CCATCCACTGTTCCCCTGT
Nono	GCCAGAATGAAGGCTTGACTAT	TATCAGGGGGAAGATTGCCCA
Ppia	GAGCTGTTTGCAGACAAAGTTC	CCCTGGCACATGAATCCTGG
C1ql1	GGGGCAACAGCAACAAATAC	CCTTGGTCAGGCAATTTGAA
Carhsp1	CCCACGCATCAGACTTCTGTA	GTAGGCAGAGGGCTAGGGA
Chrna9	CGTGTGATCTCCACCAGTGT	TCCTTCATCCCTTTATCCTTGA
Chrna10	AGCCCTTCTGCATCACGTAG	AAAGCGGTCCATTACTCTGG
Clu	CCATCTGCAACTAGCTGTGAG	TCCGTTTTCTTCGGAAGTAAGAC
Dmn3	AACTTCACATCAACGCGACC	CTCGCACTGGAGTCTCTGAT
Isl1	ATGATGGTGGTTTACAGGCTAAC	TCGATGCTACTTCACTGCCAG
Kcnq4	GCTCAACCAAGTCCGCTTC	GTGATCGGAATCAGCCACAGT
Lbh	CTGCTCTGACTATCTGAGATCGG	CGGTCAAAGTCTGATGGGTCC
Lmo7	TTGAAACAACGGATTTTCGAGC	GACGCCAGGTTTGAGCTTATT
Lmod3	TTCAAGATGGGCAGCTAGAAAAT	TCGTCAGCTCGGAGATAGGAA
Six2	CACCTCCACAAGAATGAAAGCG	CTCCGCCTCGATGTAGTGC
Slc26a5	CACTCATTATGGGAGCGAGA	TCCGTCTACTTCTGCATCCAC
Strip2	ACTGGGGTCCTGAGTCAAAG	TTGACAAAAGTGAGCAGGGC

Table 1. Sequences of oligonucleotide primers for q-PCR.

Organism	Cell Type	Replicate	Analysis type	Sample ID	Raw data accession number
Mus musculus	1,000 IHCs	Biological Replicate 1, Technical replicate 1	RNA-Sequencing	IHC_1	SRX3757335
Mus musculus	1,000 IHCs	Biological Replicate 1, Technical replicate 2	RNA-Sequencing	IHC_2	SRX3757336
Mus musculus	1,000 IHCs	Biological Replicate 2, Technical replicate 1	RNA-Sequencing	IHC_3	SRX3757337
Mus musculus	1,000 IHCs	Biological Replicate 2, Technical replicate 2	RNA-Sequencing	IHC_4	SRX3757338
Mus musculus	1,000 OHCs	Biological Replicate 1, Technical replicate 1	RNA-Sequencing	OHC_1	SRX3757339
Mus musculus	1,000 OHCs	Biological Replicate 1, Technical replicate 2	RNA-Sequencing	OHC_2	SRX3757340
Mus musculus	1,000 OHCs	Biological Replicate 2, Technical replicate 1	RNA-Sequencing	OHC_3	SRX3757341
Mus musculus	1,000 OHCs	Biological Replicate 2, Technical replicate 2	RNA-Sequencing	OHC_4	SRX3757343
Mus musculus	1,000 OHCs	Biological Replicate 3, Technical replicate 1	RNA-Sequencing	OHC_5	SRX3757344
Mus musculus	1,000 OHCs	Biological Replicate 3, Technical replicate 2	RNA-Sequencing	OHC_6	SRX3757345

Table 2. **Accession numbers for each biological sample.**

Immunocytochemistry

Cochleae were perfused with 4% formaldehyde in phosphate buffered saline (PBS) and the basilar membrane together with the organ of Corti was dissected out. The tissue was treated with 0.2% Triton X-100/PBS and goat serum (10%) was used to block nonspecific binding. The tissue was then incubated with antibodies (anti-Slc7a14: HPA045929, Lot: R43519 from Atlas Antibodies; anti-Dnm3: AB3458, Lot: GR154711-3 from Abcam) and washed with PBS, followed by incubation with secondary antibodies (Life Technologies, Lot# 1579044). The samples were mounted on glass slides with antifade solution (Prolong Antifade Kit, Invitrogen, Carlsbad, CA) before imaging on a Leica Confocal Microscope (Leica TCS SP8 MP). Three cochleae from three adult CBA/J mice were used for immunodetection.

Data Records

Raw fastq sequencing files, comprised of 2 biological repeats of IHCs and 3 biological repeats of OHCs, each with 2 technical repeats, have been deposited in the NCBI Sequence Read Archive (Data Citation 2). The individual accession numbers for each biological and technical replicate is provided in Table 2. An excel file containing the RPKM gene expression values of each biological and technical repeat of IHCs and OHCs is included as "GSE111348_Inner_and_outer_ hair_cells_RPKM.xlsx" (Data Citation 3). Since microarray-based transcriptomes of IHCs and OHCs from adult CBA/J mice are available from our previous study (Data Citation 4)[9], we aligned the expression values of all the genes detected from RNA-seq and microarray, according to the Ensembl annotated gene names (symbols). We also obtained transcriptome datasets (Data Citation 5 and Data Citation 6) of neonatal cochlear hair cells from two published studies[12,13]. The gene expression values together with transcriptome datasets from these published studies are included for comparison in Data Citation 7. Alignment of each gene from different studies was also assisted by reference to Ensembl, HGNC, Entrez Gene and OMIM. Additional resources such as the gEAR (https://www.umgear.org) and SHIELD (https://shield.hms.harvard.edu/index.html) were also used for reference and verification.

Technical Validation

RNA quality control and RNA-seq quality control

We analyzed RNA quality and concentration of our samples to determine their suitability for RNA-sequencing using an Agilent 2100 BioAnalyzer. In addition to using the 2:1 ratio (28S:18S) as an indication for determining the integrity of RNA in the electropherogram, we also used the RIN (RNA integrity number) software algorithm to evaluate the quality of our RNA samples. All of our samples had a RIN of 9, indicating that the integrity of RNA samples was high with minimal degradation.

Sequencing accuracy

We used the FastQC app (version 1.0.0) on the Illumina cloud computing interface (https://basespace.illumina.com/ome/index) to examine the quality of the reads. The analysis compared the read signals to the probability of accurate base-reading with a Phred quality score[20]. The fastq files generated from RNA-sequencing were analyzed for base-reading accuracy. All our sequencing runs exceeded 30, which reflects a 99.9% accuracy of the correct base at a given nucleotide in the sequence. This suggests that the RNA-sequencing performed was of high quality and unambiguous. We used Phred quality score $\geq$ 30 as the high-quality cutoff in our analysis for all samples.

Reproducibility of biological samples

Correlation coefficient was used to examine reproducibility of biological and technical replicates of IHCs and OHCs. Fig. 2a presents three plots of comparison among three biological replicates OHCs while Fig. 2b shows the comparison among three technical repeats of the dataset for OHCs. As shown, the data points are all concentrated near the line (replicate 1) with small deviation. The mean correlation
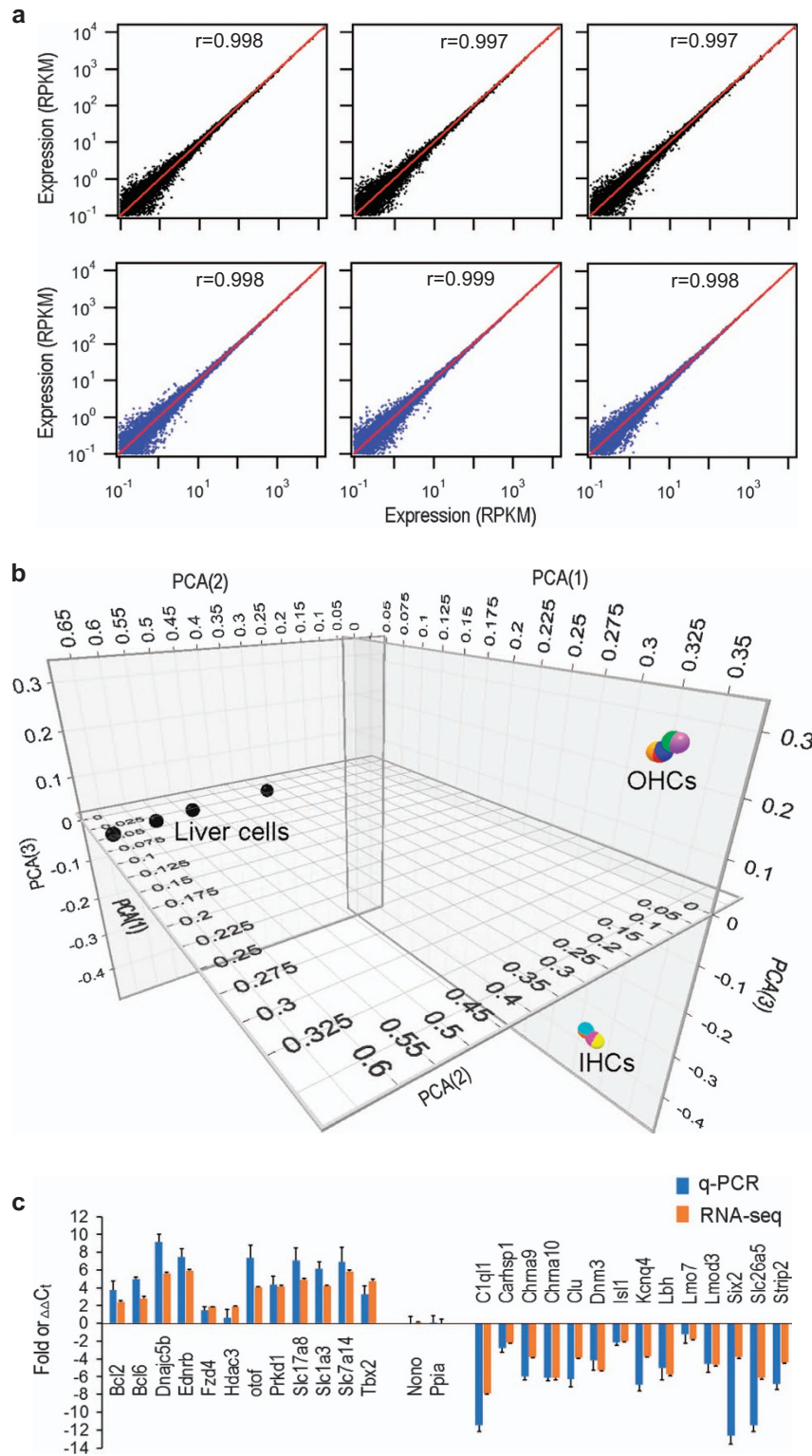
Figure 2. Reproducibility of biological replicates and RT qPCR validation of differential expression of 27 genes in IHCs and OHCs. (**a**) Correlation coefficient between biological replicates from OHCs. Correlation coefficient (r) for each comparison is given in each panel. (**b**) Correlation coefficient between technical repeats of RNA-seq from OHC samples. (**c**) PCA analysis of the gene expression profiles of IHCs and OHCs compared with liver cells. Filled circles with different colors represent different biological and technical repeats for IHCs and OHCs. (**d**) Validation of differential expression of 26 genes between IHCs and OHCs using RT qPCR and RNA-seq. Positive values indicate higher expression of the genes in IHCs than in OHCs, while negative values indicate higher expression the genes in OHCs than in IHCs. Fold differences, based from RPKM values from RNA-seq, are all calculated in log2 base. $\Delta\Delta$Ct for each gene was calculated from RT qPCR.
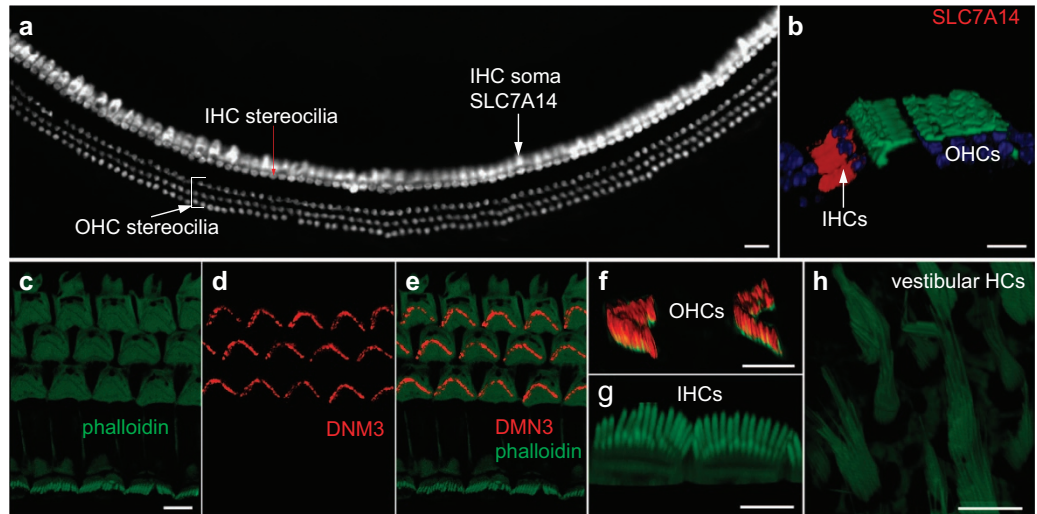
Figure 3. Differential expression of SLC7A14 and DNM3 in IHCs and OHCs by immunocytochemistry and confocal microscopy. (**a**) Expression of SLC7A14 in IHCs in a whole mount preparation from an adult mouse cochlea. Stereocilia bundles were labelled with rhodamine-phalloidin while SLC7A14 was labelled with anti-SLC7A14 antibody. Bar: 20 μm. (**b**) Confocal optical sectioning from the whole mount preparation in (**a**). Expression of SLC7A14 (in red) is only detected in IHCs. Bar: 10 μm. (**c**) Top view of stereocilia bundles of hair cells. Rhodamine-phalloidin (in green) was used to label actin. Bar: 5 μm (for c, d and e). (**d**) Expression of DNM3 (in red) in the same view as in **c**. (**e**) A merged image of c and d. (**f**): Optical sectioning image of OHC stereocilia bundles under high magnification. (**g**) Optical sectioning image of IHC stereocilia bundles. (**h**) Image of vestibular hair cells. In f, g and h, antibodies against phalloidin and DNM3 were used. Expression of DNM3 was only seen in the stereocilia bundle of OHCs. Bar: 5 μm for f and g and h.

coefficient between biological replicates of OHCs is 0.994 ± 0.0003 (mean ± SD), while the mean correlation coefficient between technical repeats of OHCs is 0.999 ± 0.0045. The correlation coefficient between biological replicates of IHCs is 0.9984 ± 0.0003 (mean ± SD), and the coefficient between technical repeats is 0.994 ± 0.0045. The analysis suggests that the results were highly reproducible.

Principal component analysis (PCA) is a technique commonly used to measure levels of variation and similarity among gene expression datasets. We used PCA to examine similarity of gene expression of different cell populations as well as reproducibility of biological replicates. Fig. 3c shows PCA of the gene expression profiles of IHCs and OHCs. Transcriptome data of mouse liver cells from a published study[21] was downloaded and normalized with our data set. As shown, the expression profiles of OHCs are highly reproducible as the data points from three biological and three technical repeats are clustered all together with small variability. Similarly, the expression profiles of IHCs are also highly reproducible. However, the datasets of IHCs and OHCs are separated by a large distance, suggesting that their gene expression profiles are different. The gene expression profile of liver cells is also distinct from those of IHCs and OHCs, as liver cells are further away from hair cells in the graph.

Real-time qPCR validation

Fifteen additional CBA/J mice were used to prepare three biological replicates of IHCs and OHCs for RT qPCR to validate the expression of 26 genes, 14 of which were highly expressed in OHCs and 12 were highly expressed in IHCs. The expression values were all normalized to the cycle threshold (Ct) value of *Nono* and *Ppia*. *Nono* and *Ppia*, used as reference genes in a previous study[19], had similar level of expression with no statistical significance between the two populations of hair cells in both previous microarray[9] and present RNA-seq studies. We compared the patterns of differential expression of these genes between IHCs and OHCs using expression values from qPCR and RNA-seq. While log2 fold difference for each gene was computed using the RPKM values of IHCs vs. OHCs from RNA-seq, the ΔΔCt for each gene was calculated from RT qPCR. Fig. 2d shows such a comparison after the expression values were normalized to fold changes. Although the values from two analyses are different, the trend of differential expression of these genes is highly consistent between the two datasets.

Immunocytochemistry

We used immunocytochemistry to detect the expression of SLC7A14 and DNM3; the function of these proteins in the two populations of hair cells has not been characterized. *Slc7a14* and *Dnm3* are

differentially expressed in IHCs and OHCs, respectively, as shown in our previous microarray-based transcriptome analysis[9]. Current study (Fig. 2d) also shows that *Slc7a14* and *Dnm3* are preferentially expressed in IHCs and OHCs, respectively. *Slc7a14* is predicted to encode a glycosylated, cationic amino acid transporter protein to mediate lysosomal uptake of cationic amino acids. This gene is expressed in the photoreceptor layer of the retina and mutations in this gene are associated with autosomal recessive retinitis pigmentosa[22]. *Dnm3* encodes dynamin-3, which is predicated to be involved in producing microtubule bundles and able to bind and hydrolyze GTP. We used antibodies (against SLC7A14 and DNM3) and confocal microscopy to determine where they are expressed and whether they are differentially expressed. As shown in Fig. 3a,b, strong staining of SLC7A14 is detected in the soma of IHCs and but not in the soma of OHCs. Thus, SLC7A14 may be used as a specific marker for IHCs. Conversely, DNM3 expression is detected in the stereocilia bundle of OHCs, but not in the bundle of IHCs and vestibular hair cells (Fig. 3c–h) suggesting that DNM3 may play an important role in the biological property of OHC stereocilia and the components of the IHC and OHC stereocilia may be different. The functional roles of these two proteins in OHCs and IHCs are yet to be determined.

Validation by comparison with published studies

Previous studies have identified and characterized many genes expressed in hair cells in developing and adult animals using immunocytochemistry, molecular biology, and electrophysiology techniques. These genes encode some proteins for unique structure and function of hair cells as well as transcription factors important for hair cell differentiation, specification and maintenance. Since the expression of these genes has been validated by either *in situ* hybridization, antibody staining or molecular biology and electrophysiology techniques, comparison of the genes detected in our RNA-seq analysis with the genes that are already be described in the inner ear is a good way to validate our dataset. We compiled a list of genes that were identified in previous studies and presented in Table 3 (available online only). In the table, the expression (RPKM) values from our RNA-seq analysis are included for comparison. As shown, most genes that were previously detected in hair cells are also expressed in our dataset. We should point out though, some genes (especially those encoding transcription factors) are known to be expressed during development and significantly downregulated in adulthood. This may explain why some genes are expressed at lower levels (e.g., *Atoh1* and *Jag2*) or no longer detected (*Foxj1*, *Scn11a*, and *Tmc2*) in adult hair cells.

Several previous studies used microarray and RNA-seq to examine the gene expression profiles of cochlear and vestibular hair cells from embryonic and neonatal mice [9,11–14] as well as hair cells in the inner ear and lateral lines of larval and adult zebrafish[11,23–27]. Comparison of our dataset with the transcriptome datasets from previous studies offers another way to validate our results. The gene names and their expression values from microarray-based transcriptomes of IHCs and OHCs from adult mice[9] are presented in Data Citation 7. Although the expression values are not directly comparable because of the two different techniques used, the majority of the genes that are detected in hair cells in different datasets are highly consistent. In the same file, we also included transcriptome datasets from neonatal mouse hair cells[12,13]. Since these datasets were obtained from a mixed population of both IHCs and OHCs from neonatal mice, some differences between the datasets are expected.

Although the expression values from microarray and RNA-seq are not directly comparable, we expect that the genes that are differentially expressed in one cell population in the two studies should largely be consistent. We used the top differentially expressed genes in IHCs and OHCs from Fig. 4b,c of the microarray study[9] for comparison. We computed the log2 fold difference between the two hair cell types from each study and present side-by-side comparison of the fold difference values from the two techniques in Fig. 4. As shown, none of the differentially expressed genes in IHCs (IHCs/OHCs in Fig. 4a) or OHCs (OHCs/IHCs in Fig. 4b) displace fold changes in the opposite direction from the two studies, suggesting that the differentially expressed genes identified by the two techniques are highly consistent. These differentially expressed genes may provide valuable information to understand different biological properties (such as structural and functional differences) of IHCs and OHCs in the adult inner ear.

Usage notes

While acquired deafness associated with age or noise exposure is more common than genetic deafness by roughly two orders of magnitude, congenital deafness occurs in 1 out of every 1,000 to 2,000 births. Hereditary hearing loss and deafness can be regarded as syndromic or non-syndromic. Mutations or deficiencies affecting approximately 140 genes have been linked to inherited syndromic or non-syndromic hearing loss[16]. Although majority of these genes are known to be expressed in the inner ear, it is important to determine whether they are expressed in hair cells. We analyzed the expression of 125 known deafness genes. Table 4 (available online only) shows expression levels of the 125 deafness genes in adult IHCs and OHCs. As shown, most of these genes are detected in hair cells. We should point out that several genes are known to be expressed during development and significantly downregulated in adulthood. Other genes may be expressed in spiral ganglion neurons, supporting cells, and stria vascularis and play important roles in those cells. Thus, it is not surprising that the expression of some genes is not
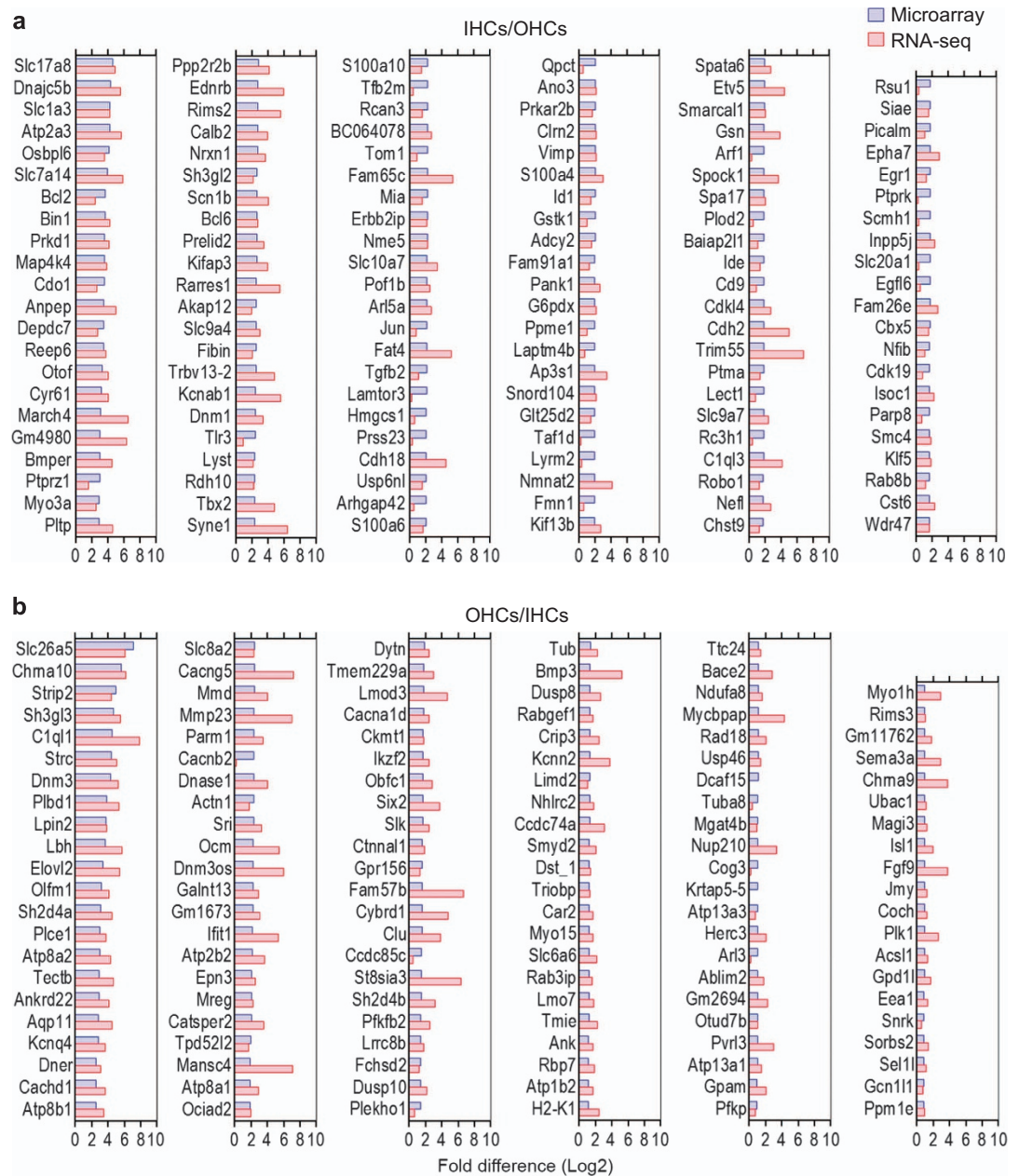
Figure 4. Comparison of differentially expressed genes in IHCs and OHCs quantified by microarray and RNA-seq techniques. The microarray data are based on genes presented in Fig. 6 of Liu *et al.*[9] (**a**) Log2 fold difference between IHCs and OHCs (IHCs/OHCs). (**b**) Log2 fold difference between OHCs and IHCs (OHCs/IHCs).

detected in hair cells. However, the analysis will be highly useful for assisting the auditory research community in exploring the function of these deafness-related genes in hair cells.

References

1. Hudspeth, A. J. Integrating the active process of hair cells with cochlear function. *Nat. Rev. Neurosci.* **15**, 600–614 (2014).
2. Fettiplace, R. & Kim, K. X. The physiology of mechanoelectrical transduction channels in hearing. *Physiol. Rev.* **94**, 951–986 (2014).
3. Dallos, P. The active cochlea. *J. Neurosci.* **12**(12), 4575–4585 (1992).
4. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**(5), 511–515 (2010).
5. Wu, Z., Wang, X. & Zhang, X. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics* **27**(4), 502–508 (2011).
6. Pan, Q, Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**(12), 1413–1415 (2008).

7. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**(7221), 470–476 (2008).

8. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**(1), 57–63 (2009).

9. Liu, H., Pecka, J. L., Zhang, Q., Soukup, G. A., Beisel, K. W. & He, D. Z. Characterization of transcriptomes of cochlear inner and outer hair cells. *J. Neurosci.* **34**, 11085–11095 (2014).

10. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**(10), 1135–1145 (2008).

11. Elkon, R. *et al.* RFX transcription factors are essential for hearing in mice. *Nat. Commun.* **6**, 8549 (2015).

12. Cai, T., Jen, H. I., Kang, H., Klisch, T. J., Zoghbi, H. Y. & Groves, A. K. Characterization of the transcriptome of nascent hair cells and identification of direct targets of the Atoh1 transcription factor. *J. Neurosci.* **35**, 5870–5883 (2015).

13. Scheffer, D. I., Shen, J., Corey, D. P. & Chen, Z. Y. Gene expression by mouse inner ear hair cells during development. *J. Neurosci.* **35**, 6366–6380 (2015).

14. Burns, J. C., Kelly, M. C., Hoa, M., Morell, R. J. & Kelley, M. W. Single-cell RNA-Seq resolves cellular complexity in sensory organs from the neonatal inner ear. *Nat. Commun.* **6**, 8557 (2015).

15. He, D. Z., Zheng, J., Edge, R. & Dallos, P. Isolation of cochlear inner hair cells. *Hear. Res.* **145**, 156–160 (2000).

16. Shearer, A. E., Hildebrand, M. S. & Smith, R. J. H. *Hereditary Hearing Loss and Deafness Overview*. University of Washington, (2017).

17. Anders, S., Theodor, Pyl, P. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

18. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery*Genome Biol.* **4**, P3 (2003).

19. Li, Y. *et al.* Transcription factors expressed in mouse cochlear inner and outer hair cells. *PLoS One* **11**(3), e0151291 (2016).

20. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).

21. Fradejas-Villar, N. *et al.* The RNA-binding protein Secisbp2 differentially modulates UGA codon reassignment and RNA decay. *Nucleic Acids Res* **45**, 4094–4107 (2017).

22. Jin, Z. B. *et al.* SLC7A14 linked to autosomal recessive retinitis pigmentosa. *Nat. Commun.* **5**, 3517 (2014).

23. Mcdermott, B. M., Baucom, J. M. & Hudspeth, A. J. Analysis and functional evaluation of the hair-cell transcriptome. *Proc. Natl. Acad. Sci. USA* **104**, 11820–11825 (2007).

24. Steiner, A. B., Kim, T., Cabot, V. & Hudspeth, A. J. Dynamic gene expression by putative hair-cell progenitors during regeneration in the zebrafish lateral line. *Proc. Natl. Acad. Sci. USA* **111**, E1393–E1401 (2014).

25. Jiang, L., Romero-Carvajal, A., Haug, J. S., Seidel, C. W. & Piotrowski, T. Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. *Proc. Natl Acad. Sci* **111**, E1383–E1392 (2014).

26. Erickson, T. & Nicolson, T. Identification of sensory hair-cell transcripts by thiouracil-tagging in zebrafish. *BMC Genomics* **16**, 842 (2015).

27. Barta, C. L. *et al.* RNA-seq transcriptomic analysis of adult zebrafish inner ear hair cells. *Sci. Data* **5**, 180005 (2018).

28. Bermingham, N. A. *et al.* Math1: an essential gene for the generation of inner ear hair cells. *Science* **284**, 1837–1841 (1999).

29. Ding, D. *et al.* Cisplatin ototoxicity in rat cochlear organotypic cultures. *Hear. Res.* **282**, 196–203 (2011).

30. Li, S., Price, S. M., Cahill, H., Ryugo, D. K., Shen, M. M. & Xiang, M. Hearing loss caused by progressive degeneration of cochlear hair cells in mice deficient for the Barhl1 homeobox gene. *Development* **129**, 3523–3532 (2002).

31. Ikoski, J., Pirvola, U., Moshnyakov, M., Palgi, J., Arumäe, U. & Saarma, M. Expression patterns of neurotrophin and their receptor mRNAs in the rat inner ear. *Hear. Res.* **65**, 69–78 (1993).

32. Neef, J. *et al.* The Ca2 + channel subunit β2 regulates Ca2 + channel abundance and function in inner hair cells and is required for hearing. *J. Neurosci.* **29**, 10730–10740 (2009).

33. Dechesne, C. J. *et al.* Identification and ultrastructural localization of a calretinin-like calcium-binding protein (protein 10) in the guinea pig and rat inner ear. *Brain Res.* **560**, 139–148 (1991).

34. Walker, R., G., Hudspeth, A. J. & Gillespie, P. G. Calmodulin and calmodulin-binding proteins in hair bundles. *Proc. Natl. Acad. Sci. USA* **90**, 2807–2811 (1993).

35. Siemens, J. *et al.* Cadherin 23 is a component of the tip link in hair-cell stereocilia. *Nature* **428**(6986), 950–955 (2004).

36. Shima, Y. *et al.* Differential expression of the seven-pass transmembrane cadherin genes Celsr1–3 and distribution of the Celsr2 protein during mouse development. *Dev. Dyn.* **223**, 321–332 (2002).

37. Maison, S. F. *et al.* Muscarinic signaling in the cochlea: presynaptic and postsynaptic effects on efferent feedback and afferent excitability. *J. Neurosci.* **30**, 6751–6762 (2010).

38. Elgoyhen, A. B., Johnson, D. S., Boulter, J., Vetter, D. E. & Heinemann, S. Alpha 9: an acetylcholine receptor with novel pharmacological properties expressed in rat cochlear hair cells. *Cell* **79**, 705–715 (1994).

39. Elgoyhen, A. B., Vetter, D. E., Katz, E., Rothlin, C. V., Heinemann, S. F. & Boulter, J. Alpha10: a determinant of nicotinic cholinergic receptor function in mammalian vestibular and cochlear mechanosensory hair cells. *Proc. Natl. Acad. Sci. USA* **98**, 3501–3506 (2001).

40. Ben-Yosef, T. *et al.* Claudin 14 knockout mice, a model for autosomal recessive deafness DFNB29, are deaf due to cochlear hair cell degeneration. *Hum. Mol. Genet.* **12**, 2049–2061 (2003).

41. Nunes, F. D. *et al.* Distinct subdomain organization and molecular composition of a tight junction with adherens junction features. *J. Cell. Sci.* **119**, 4819–4827 (2006).

42. Schmitz, F., Königstorfer, A. & Südhof, T. C. RIBEYE, a component of synaptic ribbons: a protein's journey through evolution provides insight into synaptic ribbon function. *Neuron* **28**, 857–872 (2000).

43. Morrison, A., Hodgetts, C., Gossler, A., Hrabé de Angelis, M. & Lewis, J. Expression of Delta1 and Serrate1 (Jagged1) in the mouse inner ear. *Mech. Dev.* **84**, 169–172 (1999).

44. Hartman, B. H., Hayashi, T., Nelson, B. R., Bermingham-McDonogh, O. & Reh, T. A. Dll3 is expressed in developing hair cells in the mammalian cochlea. *Dev. Dyn.* **236**, 2875–2883 (2007).

45. Yu, X., Lau, D., Ng, C. P. & Roy, S. Cilia-driven fluid flow as an epigenetic cue for otolith biomineralization on sensory hair cells of the inner ear. *Development* **138**, 487–494 (2011).

46. Shin, J. B. *et al.* The R109H variant of fascin-2, a developmentally regulated actin crosslinker in hair-cell stereocilia, underlies early-onset hearing loss of DBA/2J mice. *J. Neurosci.* **30**, 9683–9694 (2010).

47. Wallis, D. *et al.* The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival. *Development* **130**, 221–232 (2003).

48. McGee, J. *et al.* The very large G-protein-coupled receptor VLGR1: a component of the ankle link complex required for the normal development of auditory hair bundles. *J. Neurosci.* **26**, 6543–6553 (2006).

49. Odeh, H. *et al.* Mutations in Grxcr1 are the basis for inner ear dysfunction in the pirouette mouse. *Am. J. Hum. Genet.* **86**, 148–160 (2010).

50. Imtiaz, A., Kohrman, D. C. & Naz, S. A frameshift mutation in GRXCR2 causes recessively inherited hearing loss. *Hum. Mutat.* **35**, 618–624 (2014).

51. Qian, D. *et al.* Basic helix-loop-helix gene Hes6 delineates the sensory hair cell lineage in the inner ear. *Dev. Dyn.* **235**, 1689–1700 (2006).

52. Lanford, P. J. *et al.* Notch signalling pathway mediates hair cell development in mammalian cochlea. *Nat. Genet.* **21,** 289–292 (1999).

53. Kubisch, C. *et al.* KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. *Cell* **96**(3), 437–446 (1999).

54. Longo-Guess, C. M., Gagnon, L. H., Cook, S. A., Wu, J., Zheng, Q. Y. & Johnson, K. R. A missense mutation in the previously undescribed gene Tmhs underlies deafness in hurry-scurry (hscy) mice. *Proc. Natl. Acad. Sci. USA* **102,** 7894–7899 (2005).

55. Hertzano, R. *et al.* Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system. *Eur. J. Neurosci* **25,** 999–1005 (2007).

56. Deng, M., Pan, L., Xie, X. & Gan, L. Differential expression of LIM domain-only (LMO) genes in the developing mouse inner ear. *Gene Expr. Patterns* **6,** 857–863 (2006).

57. Grillet, N. *et al.* Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. *Am. J. Hum. Genet.* **85,** 328–337 (2009).

58. Di Palma, F., Belyantseva, I. A., Kim, H. J., Vogt, T. F., Kachar, B. & Noben-Trauth, K. Mutations in Mcoln3 associated with deafness and pigmentation defects in varitint-waddler (Va) mice. *Proc. Natl. Acad. Sci. USA* **99,** 14994–14999 (2002).

59. Walsh, T. *et al.* From flies' eyes to our ears: mutations in a human class III myosin cause progressive nonsyndromic hearing loss DFNB30. *Proc. Natl. Acad. Sci. USA* **99,** 7518–7523 (2002).

60. Merritt, R. C. *et al.* Myosin IIIB uses an actin-binding motif in its espin-1 cargo to reach the tips of actin protrusions. *Curr. Biol.* **22,** 320–325 (2012).

61. Avraham, K. B. *et al.* Characterization of unconventional MYO6, the human homologue of the gene responsible for deafness in Snell's waltzer mice. *Hum. Mol. Genet* **6,** 1225–1231 (1997).

62. Hasson, T., Heintzelman, M. B., Santos-Sacchi, J., Corey, D. P. & Mooseker, M. S. Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type 1B. *Proc. Natl. Acad. Sci. USA* **92,** 9815–9819 (1995).

63. Krüger, M., Schmid, T., Krüger, S., Bober, E. & Braun, T. Functional redundancy of NSCL-1 and NeuroD during development of the petrosal and vestibulocochlear ganglia. *Eur. J. Neurosci* **24,** 1581–1590 (2006).

64. Yasunaga, S. *et al.* A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nat. Genet.* **21**(4), 363–369 (1999).

65. Ahmed, Z. M. *et al.* Mutations of LRTOMT, a fusion gene with alternative reading frames, cause nonsyndromic deafness in humans. *Nat. Genet.* **40,** 1335–1340 (2008).

66. Thomas, S. *et al.* PCP4 is highly expressed in ectoderm and particularly in neuroectoderm derivatives during mouse embryogenesis. *Gene Expr. Patterns* **3,** 93–97 (2003).

67. Erkman, L. *et al.* Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development. *Nature* **381,** 603–606 (1996).

68. Goodyear, R. J. *et al.* A receptor-like inositol lipid phosphatase is required for the maturation of developing cochlear hair bundles. *J. Neurosci.* **23,** 9208–9219 (2003).

69. Pack, A. K. & Slepecky, N. B. Cytoskeletal and calcium-binding proteins in the mammalian organ of Corti: cell type-specific proteins displaying longitudinal and radial gradients. *Hear. Res.* **91,** 119–135 (1995).

70. Togashi, H. *et al.* Nectins establish a checkerboard-like cellular pattern in the auditory epithelium. *Science* **333,** 1144–1147 (2011).

71. Lai, H. C., Klisch, T. J., Roberts, R., Zoghbi, H. Y. & Johnson, J. E. *In vivo* neuronal subtype-specific targets of Atoh1 (Math1) in dorsal spinal cord. *J. Neurosci.* **31,** 10859–10871 (2011).

72. Ruel, J. *et al.* Impairment of SLC17A8 encoding vesicular glutamate transporter-3, VGLUT3, underlies nonsyndromic deafness DFNA25 and inner hair cell dysfunction in null mice. *Am. J. Hum. Genet.* **83**(2), 278–292 (2008).

73. Zheng, J. L. & Gao, W. Q. Overexpression of Math1 induces robust production of extra hair cells in postnatal rat inner ears. *Nat. Neurosci.* **3,** 580–586 (2000).

74. Yoon, H., Lee, D. J., Kim, M. H. & Bok, J. Identification of genes concordantly expressed with Atoh1 during inner ear development. *Anat. Cell Biol* **44,** 69–78 (2011).

75. Bodmer, D., Brand, Y. & Radojevic, V. Somatostatin receptor types 1 and 2 in the developing mammalian cochlea. *Dev. Neurosci.* **34,** 342–353 (2012).

76. Verpy, E. *et al.* Mutations in a new gene encoding a protein of the hair bundle cause non-syndromic deafness at the DFNB16 locus. *Nat. Genet.* **29,** 345–349 (2001).

77. Kawashima, Y. *et al.* Mechanotransduction in mouse inner ear hair cells requires transmembrane channel-like genes. *J. Clin. Invest.* **121,** 4796–4809 (2011)).,.

78. Mitchem, K. L. *et al.* Mutation of the novel gene Tmie results in sensory cell defects in the inner ear of spinner, a mouse model of human hearing loss DFNB6. *Hum. Mol. Genet* **11,** 1887–1898 (2002).

79. Ahmed, Z. M. *et al.* The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15. *J. Neurosci.* **26**(26), 7022–7034 (2006).

80. Kikkawa, Y. *et al.* Mutations in a new scaffold protein Sans cause deafness in Jackson shaker mice. *Hum. Mol. Genet* **12,** 453–461 (2003).

81. Pearsall, N., Bhattacharya, G., Wisecarver, J., Adams, J., Cosgrove, D. & Kimberling, W. Usherin expression is highly conserved in mouse and human tissues. *Hear. Res.* **174,** 55–63 (2002).

Data Citations

1. Barta, C.L. *et al. figshare* http://doi.org/10.6084/m9.figshare.5625313 (2017).
2. *NCBI Sequence Read Archive* SRP133880 (2018).
3. *Gene Expression Omnibus* GSE111348 (2018).
4. *Gene Expression Omnibus* GSE56866 (2014).
5. *Gene Expression Omnibus* GSE65633 (2015).
6. *Gene Expression Omnibus* GSE60019 (2015).
7. Li, Y. *et al. figshare* https://doi.org/10.6084/m9.figshare.6980180 (2018).

Acknowledgements

publication's contents are the sole responsibility of the authors and do not necessarily represent the official views of the NIH or NIGMS.

Author Contributions

D.Z.H. designed the experiments. Y.L., H.L., and L.C. performed the experiments. K.P.G. and K.W.B. analyzed the data. Y.L. and D.Z.H. wrote the manuscript.

Additional Information

Tables 3 & 4 are only available in the online version of this paper.

Competing interests: The authors declare no competing interests.

How to cite this article: Li, Y. *et al.* Transcriptomes of cochlear inner and outer hair cells from adult mice. *Sci. Data.* 5:180199 doi: 10.1038/sdata.2018.199 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.