scientific data



OPEN

DATA DESCRIPTOR

# Large-scale annotated dataset for cochlear hair cell detection and classification

Christopher J. Buswinka *et al.*#

**Our sense of hearing is mediated by cochlear hair cells, of which there are two types organized in one row of inner hair cells and three rows of outer hair cells. Each cochlea contains 5–15 thousand terminally differentiated hair cells, and their survival is essential for hearing as they do not regenerate after insult. It is often desirable in hearing research to quantify the number of hair cells within cochlear samples, in both pathological conditions, and in response to treatment. Machine learning can be used to automate the quantification process but requires a vast and diverse dataset for effective training. In this study, we present a large collection of annotated cochlear hair-cell datasets, labeled with commonly used hair-cell markers and imaged using various fluorescence microscopy techniques. The collection includes samples from mouse, rat, guinea pig, pig, primate, and human cochlear tissue, from normal conditions and following *in-vivo* and *in-vitro* ototoxic drug application. The dataset includes over 107,000 hair cells which have been identified and annotated as either inner or outer hair cells. This dataset is the result of a collaborative effort from multiple laboratories and has been carefully curated to represent a variety of imaging techniques. With suggested usage parameters and a well-described annotation procedure, this collection can facilitate the development of generalizable cochlear hair-cell detection models or serve as a starting point for fine-tuning models for other analysis tasks. By providing this dataset, we aim to give other hearing research groups the opportunity to develop their own tools with which to analyze cochlear imaging data more fully, accurately, and with greater ease.**

## Background & Summary

The cochlea is the hearing organ of the mammalian inner ear. It contains thousands of highly specialized mechanically sensitive cells, called hair cells. In response to the mechanical force of sound, vibrations of the tympanic membrane are transmitted along the cochlea in a travelling wave, leading to the stimulation of sensory hair cells. Organized along the length of the cochlea in a highly regular lattice, hair cells are subdivided into two types: three rows of outer hair cells (OHCs), which amplify mechanical vibration and one row of inner hair cells (IHCs), which translate that vibration into neural signals[1]. The cochlea is tonotopically organized[2], with the critical frequency to which hair cells respond most strongly determined by location. Hair cells at the basal end of the cochlea are tuned to high-frequency sounds, while those at the apex are maximally sensitive to low frequencies. Hair cells are essential for sound perception; however, they are sensitive to damage or loss as a result of cochlear insult[3]. Mammalian cochlear hair cells are terminally differentiated and do not regenerate[4,5]. As such, hair cell death leads to permanent sensorineural hearing loss.

The severity of damage, and critically, where it has occurred, are required to fully characterize the extent of any cochlear trauma. Similarly, efforts to develop otoprotective therapeutics require detailed quantification of hair cell survival. Hair cells at the base of the cochlea are generally most vulnerable to damage, putting the detection of high-frequency sounds at the highest risk of loss, from both insult and aging. Similarly, tonotopic effects are also often evaluated when testing various therapeutics and insults to the cochlea. A common route of drug delivery to the inner ear is through the round window membrane, which is located at the base of the cochlea[6]. This creates a gradient of therapeutic concentration and potentially, efficacy[7]. Therefore, when evaluating trauma

#A full list of authors and their affiliations appears at the end of the paper.

and/or treatment outcomes, researchers often need to assess hair cell loss on a large scale, along the cochlear spiral. However, with thousands of hair cells per cochlea, this task becomes time-consuming and is rarely performed, necessitating the development and use of automated analysis tools. Finally, although it is often sufficient to quantify each subtype of hair cell and record their position along the length of the cochlea, more sophisticated assessment and quantification might reveal additional characteristics beyond cell loss, such as hair bundle disarray[8]. To facilitate other types of analyses, such as expression level of a fluorescent reporter, it is desirable for detection tools to define cell area, in addition to presence vs absence.

Object detection is a well-studied branch of computer vision[9], and machine-learning algorithms have been successfully implemented to detect instances of hair cells, classify their type (OHC vs IHC), and to denote their area by a bounding box[10–12]. The most successful object-detection algorithms rely on deep neural networks and are often trained on a large set of examples to maximize their accuracy. This either requires the manual generation of very large annotated datasets, or the use of a model pretrained with a large, general dataset and "fine-tuning" it on a smaller subset of specialized training data[13]. Models trained on data for a variety of cell types are increasingly accessible, however, they may fail when presented with new or atypical data[14].

Hair cells are such a case and are not widely represented in any public dataset to date. Therefore, any deep learning model to detect hair cells requires *de novo* generation of a hair-cell-specific dataset. Comprehensive datasets of annotated cochlear hair cells are challenging to create as the cochlea is studied in different animal models, with a wide range of pathological conditions, and at different developmental time points, all of which can dramatically affect hair cell appearance. Hair cells may be imaged with different light-microscopy techniques, staining procedures, labeling procedures, or sample preparation procedures. Furthermore, tissue morphology complexities are compounded when studying developmental alterations or cochlear insults resulting in unique patterns of histopathological findings[15]. The lack of available comprehensive annotated dataset hinders the development of cochlear analysis tools.

To meet this need, we have compiled and annotated an open-source, deep-learning-ready cochlear hair cell detection dataset. This collection of data is a large collaborative effort: cochlear samples were prepared, and imaging datasets collected by independent hearing research laboratories, curated to be representative of techniques across the field of auditory neuroscience (summarized in Table 1). The collection includes samples from mouse, rat, guinea pig, pig, primate, and human cochlear tissue, from normal conditions and following *in-vivo* and *in-vitro* ototoxic drug treatment. The dataset includes over 107,000 hair cells, all of which have been identified and annotated as either inner or outer hair cells. Each image was annotated by experts (Fig. 1), and carefully validated to ensure accuracy. We further provide recommendations and exemplar code for data pre-processing, strategies for proper train-test splitting, sampling procedures for whole cochlear datasets, and processes necessary for the fine tuning of future deep-learning models. This dataset is the foundational dataset for the Hair Cell Analysis Toolbox, a deep learning-based software for cochlear analysis we previously reported[10]. It has therefore been demonstrated as sufficient to train accurate detection models. It is our hope that this dataset can be used by other research groups to train and benchmark automated hair cell detection and classification algorithms, and thereby aid in their future development.

## Methods

**Sample preparation and imaging.** *Indzhykulian laboratory.* All experiments were carried out in compliance with ethical regulations and approved by the Institutional Animal Care and Use Committee (IACUC) of Massachusetts Eye and Ear. Early postnatal day 5–7 (P5-P7) wild-type *C57Bl6* murine cochleae were dissected as a single piece in Leibovitz's L-15 culture medium (21083-027, Thermo Fisher Scientific) then fixed in 4% formaldehyde for 1 hour. Some animals used in this study were injected at P0 with AAV9-PHP.B driving eGFP expression via round window membrane. In preparation for immunolabeling, samples were then permeabilized with 0.2% Triton-X for 30 minutes then blocked with 10% goat serum in calcium-free Hanks' Balanced Salt Solution (HBSS) for two hours. Hair cells were labeled with an anti-Myosin-VIIA primary antibody (#25-6790, Proteus Biosciences, 1:400) followed by application of goat anti-rabbit CF568 (Biotium) secondary antibody. Samples were additionally labeled with Phalloidin, allowing visualization of actin filaments (Biotium CF640R Phalloidin) and DAPI to visualize the nuclei. Stained samples were mounted on a slide using ProLong Diamond Antifade Mounting kit (P36965, Thermo Fisher Scientific).

Cochlear samples were imaged with a Leica SP8 confocal microscope (Leica Microsystems) using a 63×, 1.3 NA objective. Multiple confocal Z-stacks of 512 × 512 pixels in X and Y, with an effective pixel size of 288 nm, and a Z step of 1 μm were collected using the tiling function of the Leica LASX acquisition software to encompass whole cochleae. Stacks were maximum intensity Z-projected to form 2D images and annotated as outlined below.

*Ballesteros laboratory.* Animal care and experimental procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and were approved by the Animal Care and Use Committee of the National Institute of Deafness and Other Communication Disorders (ASP1617). Early postnatal P1 or P6 murine temporal bones were isolated and fixed in 4% paraformaldehyde (PFA) in PBS for 20 min at room temperature and under gentle shaking. After fixation, organs of Corti were dissected in PBS, followed by permeabilization with 0.5% Triton-X in PBS for 30 minutes at room temperature. Permeabilized tissue was washed three times with PBS and then blocked with 10% Normal goat serum (50062Z, Invitrogen) for 4 hours at RT. Hair cells were labeled with a 1:200 dilution of anti-Myosin-VIIA primary antibody (#25-6790, Proteus Biosciences) in goat serum overnight at 4 °C, followed by application of a 1:1000 dilution of goat anti-rabbit Alexa Fluor 555 (A32732, Invitrogen) and a 1:400 dilution of Alexa Fluor 488 phalloidin (A12379, Thermo Fisher Scientific) for one hour

| Laboratory | Number of Images | OHC | IHC | Animal | Type of Microscopy | Treatment | Age | Labeled Proteins | |
|---|---|---|---|---|---|---|---|---|---|
| Artur A. Indzhykulian, MD, PhD | 17 | 35530 | 10099 | Mouse, WT | Confocal | None | P5-P7 | MYO7A | Actin |
| Angela Ballesteros, PhD | 7 | 10688 | 2982 | Mouse, WT | Confocal | None | P1, P6 | MYO7A | Actin |
| Jonathan E. Bird, PhD | 35 | 1944 | 540 | Mouse, WT | Confocal, spinning disk | None | P5-P7 | MYO7A | Actin |
| Lisa L. Cunningham, PhD and Katharine Fernandez, PhD | 116 | 5527 | 1857 | Mouse, WT | Confocal | Platinum Compounds | 17-23 wk | MYO7A | Actin |
| Albert S. B. Edge, PhD, and Yushi Hayashi, MD, PhD | 5 | 246 | 78 | Mouse, *Ndp* KO, WT | Confocal | None | P3, 2 mo | MYO7A | ESPN |
| Ksenia Gnedeva, PhD | 2 | 237 | 76 | Mouse, WT | Confocal | None | P1-P2 | MYO7A,or PV | Actin |
| Lukas D. Landegger, MD, PhD | 5 | 355 | 132 | Pig | Confocal | None | 5 mo | MYO7A | Actin |
| M. Charles Liberman, PhD | 205 | 5561 | 1787 | Human, Primate, Rat | Confocal | None | Adult | MYO7A | ESPN |
| Uri Manor, PhD | 14 | 2760 | 791 | Mouse, WT | Confocal, spinning disk | None | P5-P7 | MYO7A | Actin |
| Mark A. Rutherford, PhD, and Michael R. Deans, PhD | 6 | 157 | 80 | Mouse, WT | Confocal | None | P30 | MYO7A | Actin |
| Anthony J. Ricci, PhD | 6 | 311 | 118 | Mouse, WT | Confocal | Gentamicin | 4-8 wk | MYO7A | Actin |
| Guy P. Richardson, PhD and Corné J. Kros, MD, PhD | 38 | 2430 | 1020 | Mouse, WT | Epifluorescence | Aminoglycosides | P2 + 3div | MYO7A | Actin |
| Brikha R. Shrestha, PhD | 5 | 179 | 56 | Mouse, WT | Confocal | None | P18 | MYO7A | Actin |
| Ruben Stepanyan, PhD, and Martín L. Basch, PhD | 38 | 4502 | 1231 | Mouse, WT | Confocal, Structure Illumination | None | P5-P7 | MYO7A | Actin |
| Basile Tarchini, PhD | 8 | 291 | 94 | Mouse, WT | Confocal | None | P21-P22 | MYO7A | Actin |
| A. Catalina Vélez-Ortega, PhD | 9 | 306 | 115 | Mouse, WT | Confocal | Gentamicin | P4 + 1div | MYO7A | Actin |
| Bradley J. Walters, PhD | 40 | 4697 | 1750 | Guinea Pig | Confocal | None | 15-17 wk | MYO7A | Actin |
| Bo Zhao, PhD | 89 | 6966 | 2026 | Mouse, WT | Confocal, deconvolution | Cisplatin | P3 + 1div | Spectrin | Actin |
| Total | 645 | 82687 | 24832 | | | | | | |

**Table 1.** Summary of Annotated Data.

at room temperature. The tissue was then washed twice with PBS to remove the excess antibody. Stained samples were mounted on a slide using ProLong Diamond Antifade Mountant (P36965, Thermo Fisher Scientific).

Cochlear samples were imaged with an LSM980 confocal microscope equipment with an Airyscan 2 detector (Carl Zeiss), using an oil immersion alpha Plan-Apochromat 63 × /1.2 Oil Corr M27 objective (Carl Zeiss) and Immersol 518 F immersion media (ne = 1.518 (30 °C), Carl Zeiss). Multiple Airyscan 0.7 μm Z-stacks of 1356 × 1356 pixels in X and Y, with an effective pixel size of 50 nm, were collected using the Zen (Carl Zeiss) acquisition software's tiling function under Multiplex mode to encompass the whole organ of Corti epithelia. Stacks were maximum intensity Z-projected to form 2D images and annotated as outlined below.

*Bird laboratory.* Mouse cochleae (P5-P7) were prepared by the Indzhykulian laboratory (as described above) and shipped to the University of Florida. Preparations were carried out in compliance with ethical regulations and approved by the Institutional Animal Care and Use Committee of Massachusetts Eye and Ear. Coverslip mounted cochleae were imaged using an inverted microscope (Nikon Ti2-E) equipped with a 60 × oil immersion objective (CFI60 Plan Apochromat Lambda, 1.4 N.A.) and a spinning-disk confocal scanner (Yokogawa X1, 5000 RPM, 50 μm pinhole). Samples were sequentially excited using 405 nm, 488 nm, 561 nm and 640 nm laser lines, and emitted light filtered through a quad-band dichroic mirror followed by the appropriate bandpass filter ET455/50 m, ET525/36 m, ET605/52 m, ET705/72 m (Chroma) and captured using a Prime95B sCMOS camera (Teledyne Photometrics). Images from the sCMOS sensor were cropped to fit the back aperture of the X1 confocal scanhead (682 × 918 pixels) with an effective pixel size of 180 nm. Nikon Elements software was used to capture z-stacks with a step size of 0.5 μm. Maximum intensity projections were generated and annotated as described below.

*Cunningham laboratory.* The images were generated as part of a previously reported study; all procedures were approved by the NIDCD/NINDS Institutional Animal Care and Use Committee (Protocol #1327)[16]. At the end of both cyclic drug administration protocols, mice were euthanized via carbon dioxide asphyxiation followed by decapitation. Cochleae were rapidly dissected and perfused with 4% paraformaldehyde (PFA) at 4 °C through the round and oval windows and then post-fixed for 1 h at room temperature or overnight at 4 °C. Fixed tissue was decalcified in 0.5 M ethylenediaminetetraacetic acid (EDTA) for 48 h at room temperature or for up to 96 h at 4 °C.

The cochleae of ears processed for whole mounts were microdissected in 1 × phosphate-buffered saline (PBS, Sigma-Aldrich) into 5 isolated turns to be immunostained and imaged. Tissue was incubated in blocking solution (5% normal horse serum in 1× PBS and Triton X-100 (Sigma-Aldrich; 1:300) for 1 h and then rinsed for 15 min in PBS. Cochlear turns were immunostained with antibodies to (1) Myosin-VIIa (rabbit anti-Myosin-VIIA; Proteus Biosciences, Ramona, CA; 1:200) and (2) C-terminal binding protein 2 (mouse anti-CtBP2; BD Biosciences; used at 1:200) with secondary antibodies coupled to Alexa Fluors 647 (Invitrogen; used

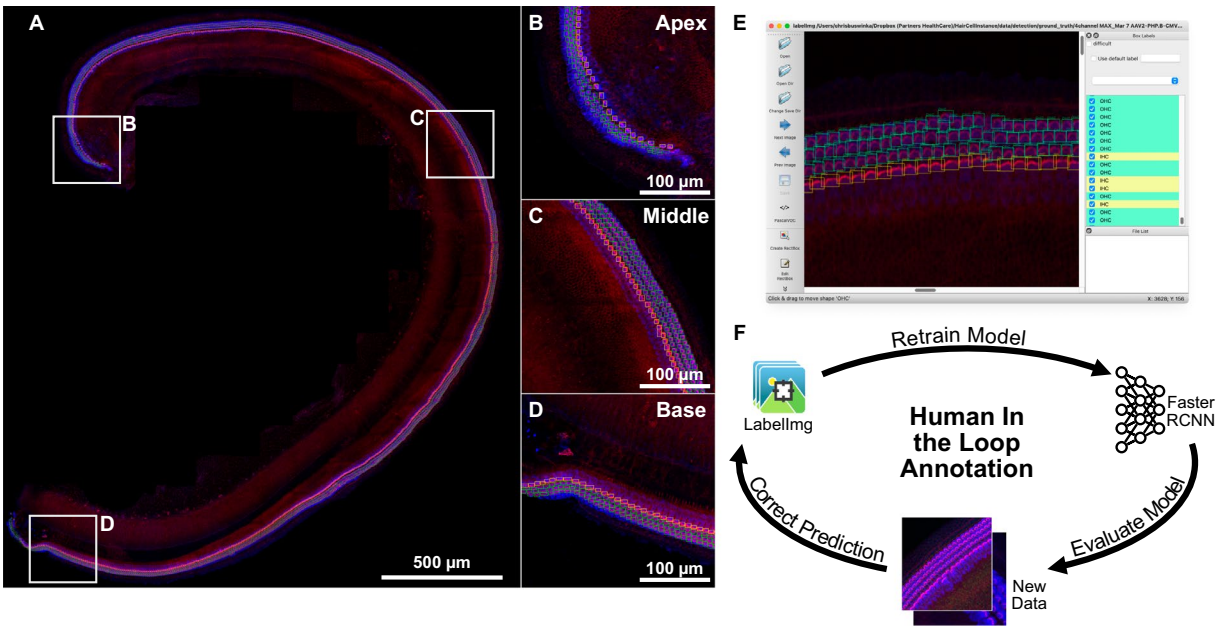


**Fig. 1** Cochlear hair cell annotation workflow using a 'human-in-the-loop' annotation paradigm. **(A)** An exemplar cochlea from our dataset with IHCs and OHCs annotated. Bounding boxes and classification labels were generated for all hair cells along the sensory epithelium. Representative regions in the **(B)** apex, **(C)** middle, and **(D)** base are shown as insets. Bounding boxes and labels were generated using labelImg **(E)**, an open-source object annotation software, with green boxes annotating OHCs and yellow boxes annotating IHCs. The annotation procedure was optimized using the human-in-the-loop paradigm **(F)**, where annotations were first generated by a preliminary neural network, then manually corrected and used to further train and improve the preliminary network, iteratively enabling more accurate candidate detections.

at 1:200) and 568 (Invitrogen; used at 1:1000), respectively. In addition, cochlear turns were stained for actin using Alexa Fluor 488-conjugated phalloidin (Invitrogen; used at 1:50) for the equimolar cyclic drug administration experiment. Due to changes in microscope settings, this antibody was used at 1:1000 for some experiments. All antibodies were diluted in 1% normal horse serum (NHS) and 30% Triton X-100. Immunostained tissue was mounted on glass slides using Fluoromount-G (Southern Biotech).

A cochlear frequency map was created via a custom plug-in for ImageJ. Confocal z-stacks (step size of 0.2 µm) corresponding to representative apical (8 kHz), middle (16 kHz), and basal (44 kHz) regions from each ear were collected using an LSM 780 laser scanning confocal microscope (Carl Zeiss AG, Oberkochen, Germany) in a 1024 × 1024 pixel raster (135 $\mu m^2$) using an oil-immersion objective (63×) of numerical aperture 1.4. IHCs and OHCs were counted in 75-µm-long stretches of the basilar membrane based on the nuclei labeled with C-terminus binding protein 2 (CtBP2). Images were then processed to generate maximum intensity projections and annotated as outlined below.

*Edge laboratory.* The images were generated as part of a previously reported study[17]. All experiments were carried out in compliance with ethical regulations and approved by the Institutional Animal Care and Use Committee of Massachusetts Eye and Ear. Cochleae were harvested from *Ndp* knockout and wild-type mice in which alkaline phosphatase was inserted into exon 2 (Stock No. 011076 Jackson Laboratory) at the ages of postnatal day 3 (P3) and 2 months (2 mo), respectively. For whole mounts of P3 mice, sensory epithelia were fixed at room temperature for 15 minutes in 4% PFA in 0.1 M phosphate buffer (pH 7.4), then rinsed with PBS. For whole mounts of 2 mo mice, temporal bones were fixed at room temperature for 2 hours in 4% PFA in 0.1 M phosphate buffer (pH 7.4), then rinsed with PBS. After decalcification with 0.12 M EDTA (pH 7.0) and dehydration with 30% sucrose, surface preparation was performed. All specimens were incubated at room temperature for 30 minutes in 10% donkey serum with 0.2% Triton X for blocking. Hair cells were labeled with an anti-Myosin-VIIA primary antibody (#25-6790, Proteus Biosciences, 1:500) followed by application of a donkey anti-rabbit AF488 secondary antibody (Invitrogen, 1:500). Samples were additionally labeled with Phalloidin, allowing visualization of actin filaments (Invitrogen AF647 Phalloidin, 1:100) and DAPI (Invitrogen, 10 µg/ml) to visualize the nuclei. Stained samples were mounted on a slide using Mounting Medium with DAPI (H-1200 VECTASHIELD, Vector Laboratories). Cochlear samples were imaged with a Leica SP8 confocal microscope (Leica Microsystems) using a 63 × , 1.3 NA objective to generate confocal Z-stacks of 512 × 512 pixels in X and Y, with an effective pixel size of 360 nm. Stacks were maximum intensity Z-projected to form 2D images and annotated as outlined below.

*Gnedeva laboratory.* All experiments were conducted according to the policies of the Institutional Animal Care and Use Committees of the Keck School of Medicine at the University of Southern California. Whole-mount

P1-2 neonatal murine cochleae were dissected in 1 × ice-cold PBS (Sigma) and fixed in 4% formaldehyde for 10 minutes. The organs were then blocked with 10% donkey serum in Tris Based Buffer supplemented with 0.1% Triton X (Sigma) and 0.01% Sodium Azide (Sigma). Hair cells were labeled with anti-Myosin-VIIA (Proteus Biosciences) or anti-Parvalbumin (Hudspeth Laboratory) primary antibodies and donkey anti-rabbit (Abcam) secondary antibodies. Phalloidin was used to visualize actin filaments (Sigma) and DAPI was used to visualize the nuclei. The samples were imaged with a Zeiss LSM 800 confocal microscope using a 20 × objective. The images were collected at 1024 × 1024 pixels with a pixel size of 0.31um.

*Landegger and Arnoldner laboratories.* All experiments were approved by the Animal Care Committee of the Medical University of Vienna and the Austrian Federal Ministry of Education, Science and Research (BMBWF-2020–0.272.252). Five-month-old pigs were euthanized at the end of cochlear implantation experiments in a prior study. Both inner ears of each animal were extracted with a hole saw and immediately stored in 4% formaldehyde (pH 7.4) for 24 hours. The contralateral, untreated side was decalcified for two months in 12% EDTA solution (pH 7.4) at 37 °C and whole mounts were dissected in 1x PBS. In preparation for immunolabeling, samples were permeabilized with 1% Triton-X and blocked simultaneously with 10% goat serum in 1× PBS for one hour. Hair cells were labeled with anti-Myosin-VIIA primary antibody (#25-6790, Proteus Biosciences, 1:200) followed by application of goat anti-rabbit 568 secondary antibody (Invitrogen, #A11011, 1:400). Samples were additionally labeled with Phalloidin for visualization of actin filaments (Alexa Fluor 488, Invitrogen #A12379, 1:500) and DAPI to mark the nuclei (1:1000). Stained samples were mounted on a side using ProLong Diamond Antifade Mounting kit (#P36930, Thermo Fisher Scientific). Cochlear samples were imaged with a Nikon Ti Eclipse confocal microscope (Nikon, Tokyo, Japan) and a 60 × 1.4NA oil immersion objective. Multiple confocal Z-stacks of 1024 × 1024 pixels in X and Y (207 nm/pix) were generated with a distance of 2.5 µm between each image scan in Z. All stacks were Z-projected at their maximum intensity to form 2D images and annotated as outlined below.

*Liberman laboratory*. Human postmortem temporal bone collection. This study was performed in line with the principles of the Declaration of Helsinki. All procedures were approved by the Mass General Brigham Institutional Review Board (IRB) overseeing human research at Mass Eye and Ear. Human-subject research in this manuscript meets the criteria for exemption from the Mass General Brigham IRB requirements in DHHS regulations (45 CFR 46), which would include participant consent. Specifically, the research falls under Exemption 4: Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. Post-mortem human inner ear tissue was obtained under IRB protocol number 2020P000508, indicating consent waiver by the IRB. Demographic data was limited to age and gender, with no protected health information included. The images were generated as part of a previous study[18]. Temporal bones were extracted at autopsy with a bone-plugging tool soon after death and immediately immersed in buffered 10% formalin after opening the round and oval windows. After post-fixation (4 °C) for at least 6 days, the bone plug containing the cochlea was drilled to remove as much of the petrous bone as possible, and then immersed in EDTA at room temperature for ~27 days. The cochlea was then microdissected into 8 – 9 pieces, each containing the osseous spiral lamina and the attached organ of Corti. Then, each piece underwent a freeze/thaw step in 30% sucrose for permeabilization, followed by 1 hr at room temperature in a blocking buffer (PBS with 5% normal horse serum and 0.3–1% Triton-X). Tissue was then incubated overnight at 37 °C with the following primary antibodies (plus 0.3–1% Triton-X): 1) rabbit anti-Myosin VI and/or VIIa (Proteus Biosciences #25-6791 and 25-6790, respectively) at 1:100 to count hair cells, 2) mouse (IgG1) anti-CtBP2 (C-terminal Binding Protein; BD Biosciences #612044) at 1:200, 3) chicken anti-neurofilament (Chemicon #AB5539) at 1:1000, and 4) goat anti-ChAT (choline acetyltransferase; Millipore #AB144P) at 1:100. Primary incubations were followed by 2 sequential 60-min incubations at 37 °C in species-appropriate secondary antibodies (coupled to Alexa Fluor dyes) with 0.3–1% Triton-X. Confocal z-stacks of the inner and outer hair cells were acquired at 14 equally spaced locations along the spiral at 240 nm/pixel in x and y (digital zoom 0.75) and with 0.33 µm z-spacing on a Leica SP8 using a 63 × glycerol objective (1.3 N.A.).

Primate inner ear tissue. The rhesus monkeys, aged 29–34 yrs, were part of an ongoing collaboration with the Ramachandran laboratory at Vanderbilt University. All experiments were carried out in compliance with ethical regulations and approved by the Institutional Animal Care and Use Committee of the Vanderbilt University, protocol numbers M1600263-00 and M1600263-01. Animals were euthanized by transcardial perfusion with 4 °C Krebs buffer (pH 7.4) followed by 4% paraformaldehyde (pH 7.4), while deeply anesthetized with sodium pentobarbital. Following euthanasia, the cochleae were exposed, the round and oval windows were punctured, and cochlear scalae were perfused with the same fixative. Cochleae were submersion-fixed for 2 h and then transferred to 0.12 M EDTA for decalcification. EDTA was refreshed weekly for 3–5 wks, and decalcified tissue was trimmed at each change. Decalcified cochleae were dissected into quarter- or half-turns, and the tissue was cryoprotected in 30% sucrose for 15 min and frozen on dry ice to permeabilize. The pieces were thawed, rinsed in PBS (pH 7.3), and incubated in a blocking reagent (5% NHS with 1% Triton-X in PBS) for 1 h at room temperature. Then, the tissue was transferred to a solution containing primary antibodies (in 1% NHS with 1% Triton-X) to label (1) pre-synaptic ribbons, with mouse (IgG1) anti-CtBP2 (C-terminal binding protein 2; BD Transduction Labs; 1:200); (2) glutamate receptor patches, with mouse (IgG2) anti-GluA2 (Millipore; 1:200) (3) hair cell cytoplasm, with rabbit anti-myo7a (Myosin-VIIA, Proteus Biosciences; 1:200); and (4) cochlear afferent and efferent fibers, with chicken anti-NFH (neurofilament-H; Chemicon; 1:1000). Following an 18 h incubation in primary antibodies at 37 °C, the tissue was rinsed in PBS and incubated in species-appropriate secondary antibodies

(coupled to AlexaFluor fluorophores) in two separate 1 h incubations. Finally, the tissue was rinsed and mounted in Vectashield (Vector Laboratories, Inc.), and confocal z-stacks were acquired with a 63× glycerol-immersion objective (N.A. = 1.3) on a Leica TCS SP8, at 140 nm/pixel in x and y (digital zoom 1.28) and with 0.33 μm z-spacing.

Rat inner ear tissue. The rats were part of an unpublished collaboration with the Salvi lab at SUNY Buffalo. Cochleae were fixed by immersion in 4% paraformaldehyde at 4 °C for 24 h, decalcified with 10% EDTA at 4 °C for 4 days, and then rinsed with 0.1 M PBS. Subsequent immunostaining protocols were similar to those described above. Confocal z-stacks were acquired with a 63× glycerol-immersion objective (N.A. = 1.3) on a Leica TCS SP8, at 70 nm/pixel in x and y (digital zoom 2.4) and with 0.33 μm z-spacing.

*Manor laboratory.* Samples were prepared as described in the Indzhykulian Laboratory procedures above out in compliance with ethical regulations and approved by the Animal Care Committee of Massachusetts Eye and Ear, then imaged with two microscope systems. For Airyscan imaging, cochlear samples were imaged with a Zeiss LSM 880 Rear Port Laser Scanning Confocal and Airyscan FAST microscope using a 63×, 1.4 NA objective. Multiple confocal Z-stacks of 1792 × 1792 pixels in X and Y, with an effective pixel size of 43 nm, and a Z step of 0.5 μm were collected using the tiling function of the Zeiss Zen Black 2.3 acquisition software to encompass representative basal, middle, and apical cochlear regions. For spinning disk imaging, cochlear samples were imaged with a Zeiss CSU Spinning Disk Confocal microscope equipped with a Yokogawa spinning disk scan head and an EM-CCD camera using a 40×, 1.3 oil objective. Multiple confocal Z-stacks of 11180 × 8944 pixels in X and Y, with an effective pixel size of 229 nm, and a Z step of 0.5 μm were collected using the tiling function of the Zeiss Zen Blue 2.3 software to encompass whole cochleae.

*Ricci laboratory.* The images were generated as part of a previously reported study[19]. All animal studies were conducted in strict accordance with the protocols approved by the Institutional Animal Care and Use Committee at Stanford University (APLAC-14345). We utilized C57BL/6 mice (Charles River: 026) of both sexes, aged between 4 and 8 weeks. Gentamicin (180 mg/kg) was systematically administered through intraperitoneal (IP) injection. Thirty minutes later, furosemide (100 mg/kg) was also injected to expedite gentamicin-induced ototoxicity. These injections were repeated daily for a period of 10 days. The cochleae were harvested from the animals and fixed with 4% PFA/PBS (15710, EMS). To decalcify the tissue, 0.5 M EDTA (E177, VWR) was applied for six hours. We dissected the whole-mount organ of Corti and removed the tectorial membrane. Subsequently, the tissue was permeabilized for 30 minutes in 0.5% Triton-X/PBS (BP151, Fisher Scientific) and then blocked for two hours in 5% BSA/PBS (BP1600, Fisher Scientific, Fair Lawn, NJ) at room temperature. For staining, we employed primary antibodies against Myosin-VIIa (hair cell marker; 1:500; MYO7A 138-1, DSHB), which were incubated overnight at 4 °C and then for an additional two hours at 37 °C in the same blocking buffer. Following three washes with the blocking buffer, the tissues were incubated with secondary antibodies conjugated with Alexa Fluor 546 (1:500; A10036, Life Technologies) and Alexa Fluor 647-conjugated phalloidin (F-actin marker; 1:200; A22287; Invitrogen) for an hour at room temperature. After five washes with PBS, the tissues were mounted on a glass slide using ProLong Diamond Antifade Mountant with DAPI (P36962, Invitrogen) and stored in the dark at 4 °C. Z-stack images of cochlear tissues were acquired using a confocal microscope (LSM700, Zeiss, Jena, Germany) equipped with a 40× objective (NA 1.3). Each individual image comprises a resolution of 1660 × 1660 pixels (X × Y) with a pixel size of 96 nm. The images were collected at intervals of 1 μm along the Z-axis. Z-stack projections using average intensity were generated using ImageJ software.

*Richardson and Kros laboratories.* The images were generated as part of a previously reported study[20] assaying hair cell survival in response to aminoglycosides and putative protective compounds from the Life Chemicals Diversity Set. Briefly, P2 mouse cochlear cultures were prepared from wild-type CD-1 mice of either sex. Organs of Corti were plated onto collagen-coated (Corning, 354236) coverslips in 93% DMEM-F12, 7% FBS, and 10 μg/mL ampicillin, and maintained for 24 hours at 37 °C in Maximow slide assemblies. This work was carried out in accordance with the Home Office Animals (Scientific Procedures) Act of 1986 and approved by the University of Sussex Animal Welfare Ethical Review Boards.

Following 24 hours of incubation, coverslips with adherent cochleae were placed in 35 mm Petri dishes (Greiner Bio-One, 627161), and 1 mL culture media (98.8% DMEM/F12, 1.2% FBS) containing either vehicle (0.5% DMSO), 5 μM gentamicin with 0.5% DMSO, or 5 μM gentamicin with 50 μM test compound UoS-7692. After 48 hours, cultures were washed in PBS, fixed at room temperature for 1 hour in 3.7% formaldehyde (v/v) (MilliporeSigma, F1635) in 0.1 M sodium phosphate pH 7.4. For labeling cultures were incubated at 4 °C overnight in PBS containing 10% horse serum and 0.1% Triton-X with 1:200 Texas red phalloidin (Invitrogen, Thermo Fisher Scientific, T7471) and 1:1000 rabbit anti-myosin VIIa (#25-6790, Proteus Biosciences), followed by 4 hours at room temperature in 1:500 Alexa Fluor 488 goat anti-rabbit secondary antibody (Invitrogen, Thermo Fisher Scientific, A-11034). Cultures were mounted in Vectashield (Vector Laboratories, H-1000) and imaged in a mid-basal region, approximately 20% along the length of the cochlea from the basal end, using a 40 × 0.75 NA objective on a Zeiss Axioplan2 upright microscope with a Spot RT Slider camera with an effective pixel size of 183 nm, then scaled to 290 nm. In order to facilitate quantification, images were obtained from multiple focal planes and merged into one image per channel in which all cells were in focus using Adobe Photoshop Creative Cloud.

*Rutherford and Deans laboratories.* All procedures were approved by the Animal Studies Committee at Washington University in St. Louis and IACUC at the University of Utah in Salt Lake City. Briefly, mice were euthanized by $CO_2$ inhalation and perfused through the heart with fixative solution (4% PFA in sodium

phosphate buffer). Temporal bones were isolated in PBS. A hole was made through the bone near the cochlear apex, the round and oval windows were punctured, and the temporal bone was bath fixed in fixative solution for one hour at 4 °C. The samples were decalcified in 10% EDTA for 3 hours. Whole-mount preparations containing organ of Corti and spiral ganglion were isolated from the cochlea in three pieces containing the apical, middle, or basal thirds. After incubation in blocking buffer containing 10% donkey serum and 0.2% Triton-X at room temperature for 2 hours, antibodies were applied in blocking buffer at 4 °C overnight: (Myosin-VIIA rabbit, Proteus Biosciences; Alexa Fluor 555 Phalloidin). The Myosin-VIIA antibody was then labeled with AlexaFluor 488 for 1 hour at room temperature. Samples were mounted on slides in Mowiol mounting medium and imaged on a Zeiss LSM700 with a z-step of 0.5 micrometers and pixel size of 100 nm in x and y with a 63 × 1.4NA oil objective lens (Carl Zeiss). Image stacks were maximum intensity Z-projected to form 2D images and annotated as described below.

*Stepanyan and Basch laboratories.* Samples were prepared as described in the Indzhykulian Laboratory procedures above in compliance with ethical regulations and approved by the Animal Care Committee of Massachusetts Eye and Ear, then imaged on Zeiss Structure Illumination Microscope system with 20 × 0.8 NA and 40 × 1.3 NA objective lenses. Zen 2.3 pro software was used for image acquisition with an Axiocam 702 mono camera (1920 × 1216 pixels). Images were collected as a Z-stack with a 1 μm step and 294 nm (20 × lens) or 147 nm (40 × lens) pixel size. Image stacks were then maximum intensity Z-projected to form 2D images and annotated as described below.

*Shrestha laboratory.* All experiments were carried out in compliance with regulations and procedures approved by the IACUC of Massachusetts Eye and Ear. Postnatal day 18 (P18) mice of both sexes were euthanized by administering an overdose of pentobarbital sodium (Fatal Plus, 150 mg/kg). Temporal bones were quickly removed and immersed for 1 hour in fresh 4% PFA in 1× PBS. They were then decalcified in 120 mM EDTA at 4 °C for 48 hours. Cochleae were dissected out and stored in 1× PBS at 4 °C until further use 24-72 hrs later. Immunostaining was performed on these whole mount preparations as follows: they were permeabilized and blocked for 1 hr in 1x PBS with 1% Triton X-100 (PBS-Tx) and 5% normal donkey serum (NDS). Tissues were then incubated overnight at 4 °C with rabbit anti-Parvalbumin (1:250, Swant #PV27a) in 0.3% PBS-Tx and 1% NDS. After tissues were rinsed in PBS-Tx for 45 mins, they were incubated for 2 hours at room temperature with Alexa Fluor 555 donkey anti-rabbit (1:500) and Alexa Fluor 488 Phalloidin (1:500, Invitrogen #A12379) in 0.3% PBS-Tx and 1% NDS. After rinsing again for 45 minutes in PBS-Tx, all tissues were mounted in Fluoromount-G (Electron Microscopy Sciences, #17984-25). Images were acquired from the middle turn at 90 nm/pixel XY resolution (1024 × 1024 pixels) and 0.33 μm optical sectioning in the Z dimension using a Leica SP8 confocal microscope fitted with a 63 × 1.3 NA objective (digital zoom factor 2). Maximum intensity projections of these image stacks were created in Fiji. The resulting images were then annotated as outlined below.

*Tarchini laboratory.* The images were generated as part of a previously reported study[21]. Animals were maintained under standard housing and all animal work was reviewed for compliance and approved by the Animal Care and Use Committee of The Jackson Laboratory. Briefly, temporal bones of postnatal day 21-22 mice were isolated, and the cochlea punctured at the apex before immersion fixation in 4% PFA for 1 h at 4 °C. Temporal bones were then incubated overnight at room temperature in 0.11 M EDTA for decalcification before exposing the auditory epithelium. Samples were blocked and permeabilized in PBS with 0.5% Triton-X and bovine serum albumin (1%) for at least 1 h at room temperature. Primary and secondary antibodies were each incubated overnight at 4 °C in PBS. After each antibody incubation, samples were washed 3 times with PBS + 0.05% Triton-X before a final post-fixation in PFA 4% for 1 h at room temperature. Antibodies used were rabbit MYO6 (#25-6791, Proteus Biosciences) and mouse MYO7 (DSHB; MYO7A 138-1). Secondary antibodies were raised in donkey and coupled to Alexa Fluor (AF) 555 or 647 (ThermoFisher Scientific; anti-rabbit AF555, A-31572; anti-mouse AF555, A-31570; anti-mouse AF647, A-31571). We used fluorescent conjugated phalloidin to reveal F-Actin (ThermoFisher Scientific: AF488, A12379). Samples were then mounted flat on a microscopy slide (Denville M1021) directly under a 18 × 18 mm #1.5 cover glass (VWR 48366-045), using Mowiol (10% w/v) (Calbiochem/MilliporeSigma 4759041) as mounting medium. Images were captured with a LSM800 line scanning confocal microscope using the Zen 2.3 or Zen 2.6 software, the Airyscan detector in regular confocal mode, and a 63 × 1.4 NA oil objective lens (Carl Zeiss AG). Confocal stacks were taken at the cochlear base, with a z-step of 0.1 μm, a pixel size of 68 nm, and fields of 1495 × 1495 pixels. Stacks were maximum intensity Z-projected to form 2D images and annotated as outlined below.

*Vélez laboratory.* All animal procedures were approved by the University of Kentucky Animal Care and Use Committee (protocol 2020-3535). Organ of Corti explants were isolated from C57BL/6 mice at postnatal day 4 in cold Leibovitz's L-15 medium (cat. # 21083027, Gibco) and held in place by two glass fibers glued to the bottom of a Petri dish. Explants were cultured in high-glucose DMEM (12430062, Gibco) supplemented with 7% FBS (16140071, Gibco) and 10 mg/L ampicillin (171254, Calbiochem) at 37 °C and 5% $CO_2$, for 24 h in the presence of 62.5 μM gentamicin (G1397, Sigma-Aldrich) or in control conditions. Next, explants were fixed in 4% PFA (15710, Electron Microscopy Sciences) for 24 h at 4 °C. Samples were then permeabilized with 0.5% Triton-X (22142, Electron Microscopy Sciences) for 1 h and blocked with 10% normal goat serum (10000C, Invitrogen) and 0.25% Triton-X for 1 h, at room temperature. Hair cells were labeled with rabbit polyclonal anti-Myosin-VIIA primary antibody (#25-6790, Proteus Biosciences, concentration 1:100) for 24 h at 4 °C followed by labeling with goat anti-rabbit Alexa fluor 488 (A11034, Invitrogen, 1:1000) as a secondary antibody for 3 h at room temperature. Samples were counterstained with 1 unit of rhodamine phalloidin for 30 min at room temperature (R415, Invitrogen), and mounted in ProLong Diamond antifade medium (P36961, Invitrogen). Imaging was

performed with a Leica SP8 upright confocal microscope equipped with a Leica HCX PL APO 100 × 1.44 NA objective lens. Confocal z-stacks were taken with a voxel size of 114 nm in X and Y, and 500 nm in Z.

*Walters laboratory.* All use of animals for these studies was approved by the Institutional Animal Care and Use Committee at the University of Mississippi Medical Center. Guinea pigs (*Cavia albino*) around the age of 15–17 weeks old were euthanized by deep anesthesia with isoflurane followed by a 200 mg/kg injection of Fatal Plus (Vortech Pharmaceuticals). After confirmatory decapitation, the temporal bones were removed, trepanned, and perfused with a solution of 4% PFA in PBS and then immersed in 4% PFA in PBS for 24 hours at 4 °C. The temporal bones were then decalcified in 0.25 M EDTA for 5 days on a paddle rotator at room temperature. The decalcified bony labyrinths were then dissected to remove the sensory epithelia, which were stored in PBS until ready for immunostaining. Free floating samples were washed several times in PBS, blocked for 1 hour in 10% normal goat serum, 1% bovine serum albumin, and 1% Triton-X in PBS, then immersed in primary antibodies in blocking buffer that was diluted by half in PBS. The antibodies used were Rabbit anti MYO7A (#25-6790, Proteus Biosciences) at 1:200, and Mouse anti-BRN3C (Santa Cruz Biotechnology, #sc-81980) at 1:250. The samples were next washed 3 × 10 minutes in PBS, then immersed in diluted blocking buffer containing Alexa-Fluor-conjugated secondary antibodies (Goat anti-Rabbit and Goat anti-Mouse IgG1, Invitrogen / Thermo-Fisher), each at 1:800 dilution, for 3 hours at room temperature. Cell nuclei in all samples were counterstained with Hoechst33258 at 1:1500 in PBS for 30 minutes and a subset of samples were simultaneously labeled by the addition of methanol reconstituted, AlexaFluor conjugated, phalloidin (Invitrogen/ThermoFisher) at 1:200 into the diluted Hoechst solution. Samples were then washed 3 × 5 minutes in PBS and mounted to slides using FluoroGel with DABCO (Electron Microscopy Sciences, #17985) and 1.5 thickness coverglass (Fisherbrand, Fisher Scientific). Z-stack images were then acquired under a 20 × objective (0.8 NA) using a Zeiss LSM880 point-scanning confocal microscope. Large fields of view were collected using an XY resolution of 2048 × 2048 and a Z-slice thickness around 5 μm and 188 nm pixel size. Images were then processed to generate maximum intensity projections and annotated as outlined below.

*Zhao laboratory.* Animal experiments were carried out in accordance with the National Institutes of Health Guide and were approved by the Institutional Animal Care and Use Committee of Indiana University School of Medicine. Cochlear explants dissected from P3-P4 mice were cultured overnight in DMEM/F12 media at 37 °C in a 5% $CO_2$ humidified atmosphere. Then, cochlear explants were exposed to saline, cisplatin, or other chemical solutions for two days. Samples were fixed in a fixative containing 4% PFA in HBSS for 20 minutes. After washing with HBSS, the tectorial membrane was removed. After blocking with HBSS containing 5% bovine serum albumin (BSA) and 0.5% Triton-X at room temperature for 20 minutes, samples were incubated with primary antibodies overnight at 4 °C. Samples were washed in HBSS and incubated with secondary antibodies for 2 hours at room temperature. Then, tissues were mounted in ProLong Antifade Reagents (ThermoFisher Sci). Stacked images were captured by a DM6 FS automated deconvolution microscope (Leica) using a 20 × objective (HC PL FLUOTAR 20x/0.55; 214 nm/pixel). Antibodies used in this study were anti-βII spectrin (1:200, cat# sc-136074, Santa Cruz) and Alexa Fluor 488 goat anti-mouse (1:2,000; cat# A11017, Life technologies Corporation).

**Annotation procedure.** To qualify for annotation, cells were required to have a clear and definable central feature, either a well-stained cuticular plate or a stereocilia bundle, connected to a labeled cell body of the expected shape and size for that tissue type/preparation. Most of the images within this collection were of hair cells with relatively normal appearance, with some instances of missing hair cells. Many of the annotated images are maximum projections of serial images acquired at multiple focal planes, often with the cytosol of one cell occluding another rendering the cytosol as an unreliable feature for bounding box detection. Thus, each hair cell annotation box bounds the cuticular plate enclosing the stereocilia bundle and is assigned a class label of cell type (IHC or OHC).

Bounding-box annotations were created and corrected in the labelImg open-source software[22] and the HCAT open source software. The annotations were generated using a "human-in-the-loop"[23] paradigm as outlined in Fig. 1. Initial candidate annotations were generated using a neural network, which were then audited and refined to ensure every cell classification label was correct, and that bounding boxes tightly enclose the cell's cuticular plate and stereocilia bundle without including any additional features. Since the images within the dataset were annotated by several observers, each image was then reviewed by a single lead annotator to ensure accuracy and minimize inter-annotator biases. Data annotations were saved as a separate xml file in the coco format with an identical filename to the associated image[24,25]. Although uncommon, some images within the dataset were collected following application of certain insults and may include hair cells with damaged stereocilia bundles.

## Data Records

The dataset[26] presented here is hosted with Zenodo[25] data repository with version 0.3.2 and is subdivided by type of microscopy, type of treatment, species and research group of origin. Collectively, we have annotated a total of **107,519** hair cells across over 645 images from over 20 research groups[26]. The dataset presents examples of auditory hair cells of six species: Mouse, Rat, Guinea Pig, Pig, Primate and Human, imaged at differing pixel sizes, magnifications, using point scanning confocal, spinning disk confocal, structured illumination, or widefield fluorescence microscopy techniques. Depending on origin, individual images may either contain the entire cochlea or smaller sub-regions. A detailed summary of the dataset, along with the relevant parameters is presented in Table 1. The materials and methods used by each group are reported within the *Methods* section of this manuscript.

As many of our annotated images are of full cochleae, we included exemplar functionality to selectively sample each image when training, allowing precise control to correct for class imbalance. In this way, an underrepresented image type may be shown to the model more often, improving overall generalizability. Finally, we

present an exemplar data augmentation procedure for our dataset, which can aid in training accurate models in the future. Detailed descriptions of these procedures were previously reported[10] and briefly summarized within the *Methods* section.

## Technical Validation

To ensure the accuracy of our dataset, we employed an iterative process whereby a set of expert annotators correct predicted candidate annotations from an existing deep-neural network. The process is as follows: 1) the detection neural network was trained on a preliminary subset of images with manual annotations, 2) the resulting network was evaluated on new images to generate candidate annotations, 3) these candidate annotations were inspected and corrected (as needed) by at minimum two trained annotators followed by 4) periodical re-training the detection network using newly generated pool of ground truth annotations and repeating the process. Priority was given to manually corrected poorly performing examples at each iterative stage, thereby allowing the network to direct the annotation efforts toward generalizability. Final candidate annotations were always corrected and approved by the lead annotator to eliminate any possible systematic differences in annotation style between annotators and ensuring high quality. In some cases, images contain examples of atypical cells, which are challenging to annotate. In these cases, researchers from the laboratory of origin were consulted on annotations. We have further validated the utility of this dataset in a previously reported study[10] and show that this dataset on its own is sufficient to generate accurate deep-learning detection models.

## Usage Notes

All imaging data along with the annotations can be found at: https://zenodo.org/records/10476880[26]. Our dataset contains images which represent vastly different spatial areas. Details of sample preparation, stain, and resolution are embedded in its annotation metadata. We present each maximum intensity projection image in the dataset in the TIFF-format, along with associated ground truth detection annotations in XML. For ease of use, we have created a small python library, which can automatically download and parse the dataset into a code-accessible format. Many of these images are too large to pass through a deep neural network. Therefore, we recommend sampling equal sized regions of each image a pre-determined number of times. Class and frequency imbalance issues can therefore be mitigated by choosing the sampling factor for each image, or each group of images based on the species, animal's age, type of insult, the laboratory of origin, etc. Furthermore, as the cochlea is a spiraling organ, much of the area in some of our larger images contains no hair cells (center of the spiral), therefore we recommend centering the sampled region by an existing cell annotation. Cochlear hair cells are tightly ordered, clearly visible in a majority of the images. While these data are sufficient to train accurate and generalizable hair cell detection networks, we do not believe there are enough samples to accurately train networks to predict the location of missing cells. While these data are principally meant for cell detection tasks, they may also be suitable for additional meta-analysis correlating underlying methodology to cell structure or morphology. However, we cannot prove this usefulness here.

Often, a plethora of data are not enough to prevent overtraining in deep neural networks. We therefore recommend extensive data augmentation when using this dataset for training. Of critical importance are affine transformations, which randomly rotate, shear, and scale the data. This can be challenging when working with bounding boxes, as they may increase in area when rotated. Thus, we recommend shrinking rotated boxes slightly when rotated. In general, augmentation of training data makes a resulting model invariant to the augmentations. Therefore, we recommend the following augmentations to improve model generalizability: 1) shuffle the order of fluorescence channels (switching colors) on an image to simulate changes in image structure (i.e., MYO7A is not always green, phalloidin is not always red, etc.), 2) adjust the brightness and contrast of each channel separately to simulate weak or strong staining, 3) add random noise to simulate low signal-to-noise ratio imaging, and 4) blur the image to reduce dependency of the model to proper focus or high resolution. A detailed description of our recommended augmentation procedure was previously reported[10] and successfully implemented in the associated data augmentation pipeline.

## Code availability

All associated code for downloading, loading, and preprocessing this dataset may be found at: https://github.com/indzhykulianlab/hcat-data.

Received: 30 October 2023; Accepted: 3 April 2024;
Published online: 23 April 2024

## References

1. Lim, D. J. Functional structure of the organ of Corti: a review. *Hearing research* **22**, 117–146 (1986).
2. Ashmore, J. Tonotopy of cochlear hair cell biophysics (excl. mechanotransduction). *Current opinion in physiology* **18**, 1–6, https://doi.org/10.1016/j.cophys.2020.06.010 (2020).
3. Schmiedt, R. A. Acoustic injury and the physiology of hearing. *J Acoust Soc Am* **76**, 1293–1317, https://doi.org/10.1121/1.391446 (1984).
4. Ryan, A. F., Kujawa, S. G., Hammill, T., Le Prell, C. & Kil, J. Temporary and Permanent Noise-induced Threshold Shifts: A Review of Basic and Clinical Observations. *Otol Neurotol* **37**, e271–275, https://doi.org/10.1097/MAO.0000000000001071 (2016).
5. Kurabi, A., Keithley, E. M., Housley, G. D., Ryan, A. F. & Wong, A. C. Cellular mechanisms of noise-induced hearing loss. *Hear Res* **349**, 129–137, https://doi.org/10.1016/j.heares.2016.11.013 (2017).
6. Jero, J. *et al*. Cochlear gene delivery through an intact round window membrane in mouse. *Human gene therapy* **12**, 539–548 (2001).
7. Salt, A. N. Dexamethasone concentration gradients along scala tympani after application to the round window membrane. *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* **29**, 401 (2008).

8. Wu, P.-Z. & Liberman, M. C. Age-related stereocilia pathology in the human cochlea. *Hearing Research* **422**, 108551, https://doi.org/10.1016/j.heares.2022.108551 (2022).
9. Zhao, Z.-Q., Zheng, P., Xu, S.-t & Wu, X. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems* **30**, 3212–3232 (2019).
10. Buswinka, C. J., Osgood, R. T., Simikyan, R. G., Rosenberg, D. B. & Indzhykulian, A. A. The hair cell analysis toolbox is a precise and fully automated pipeline for whole cochlea hair cell quantification. *Plos Biology* **21**, e3002041 (2023).
11. Urata, S. *et al*. Cellular cartography of the organ of Corti based on optical tissue clearing and machine learning. *eLife* **8**, https://doi.org/10.7554/eLife.40946 (2019).
12. Cortada, M., Sauteur, L., Lanz, M., Levano, S. & Bodmer, D. A deep learning approach to quantify auditory hair cells. *Hearing Research* **409**, 108317 (2021).
13. Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *Journal of Big data* **3**, 1–40 (2016).
14. Taylor L., & Nitschke, G. Improving Deep Learning with Generic Data Augmentation. 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 1542–1547, https://doi.org/10.1109/SSCI.2018.8628742 (2018)
15. Bommakanti, K., Iyer, J. S. & Stankovic, K. M. Cochlear histopathology in human genetic hearing loss: state of the science and future prospects. *Hearing research* **382**, 107785 (2019).
16. Gersten, B. K., Fitzgerald, T. S., Fernandez, K. A. & Cunningham, L. L. Ototoxicity and Platinum Uptake Following Cyclic Administration of Platinum-Based Chemotherapeutic Agents. *J Assoc Res Otolaryngol* **21**, 303–321, https://doi.org/10.1007/s10162-020-00759-y (2020).
17. Hayashi, Y., Chiang, H., Tian, C., Indzhykulian, A. A. & Edge, A. S. Norrie disease protein is essential for cochlear hair cell maturation. *Proceedings of the National Academy of Sciences* **118**, e2106369118 (2021).
18. Wu, P. *et al*. Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. *Neuroscience* **407**, 8–20 (2019).
19. Kim, J. & Ricci, A. J. *In vivo* real-time imaging reveals megalin as the aminoglycoside gentamicin transporter into cochlea whose inhibition is otoprotective. *Proc Natl Acad Sci USA* **119**, https://doi.org/10.1073/pnas.2117946119 (2022).
20. Kenyon, E. J. *et al*. Identification of a series of hair-cell MET channel blockers that protect against aminoglycoside-induced ototoxicity. *JCI Insight* **6**, https://doi.org/10.1172/jci.insight.145704 (2021).
21. Jarysta, A. & Tarchini, B. Multiple PDZ domain protein maintains patterning of the apical cytoskeleton in sensory hair cells. *Development* **148**, dev199549 (2021).
22. Lin, T. (Github, 2015).
23. Wu, X. *et al*. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems* (2022).
24. Lin, T.-Y. *et al*. in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. 740-755 (Springer).
25. OpenAIRE & Research, E. O. F. N. *Zenodo*, https://doi.org/10.25495/7gxk-rd71 (2013).
26. Buswinka, C. J. *et al*. Large-scale annotated dataset for cochlear hair cell detection and classification. *Zenodo* https://doi.org/10.5281/zenodo.10476880 (2024).


## Acknowledgements

We would like to extend our sincere gratitude to the laboratories that generously offered to contribute data for our study, even though, for various reasons, we were unable to incorporate it. Your willingness to collaborate and share your expertise is greatly appreciated, and we are pleased to note that your views are aligned with the most recent NIH policies and guidelines on data sharing. This work was supported by: R01DC020190, R01DC017166 and R01DC017166-04S1 to AAI; T32 DC000038 to CJB; R01DC018827 to GB and JEB; R01DC016365, R21DC020312 and N00014-18-1-2716 to BJW; ZIA DC-000079 to LLC; R01DC000188 and P50DC015857 to MCL; NIDCD DIR DC000096 to AB; R01DC021325 to ACV; R01DC014712 to MAR; DBR and UM were supported by the David F. and Margaret T. Grohne Family Foundation, Core Grant application NCI CCSG (CA014195), R01DC021075 and the CZI Imaging Scientist Award https://doi.org/10.37921/694870itnyzk from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation (funder https://doi.org/10.13039/100014989).

## Author contributions

Christopher J. Buswinka, Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Visualization, Supervision, Writing – Original Draft. David B. Rosenberg, Data Acquisition, Data Curation, Writing – Review & Editing. Rubina G. Simikyan, Data Curation, Writing – Review & Editing. Richard T. Osgood, Data Acquisition, Data Curation, Validation, Writing – Review & Editing. Katharine Fernandez, Data Acquisition, Writing-Review & Editing. Hidetomi Nitta, Data Curation, Writing – Review & Editing. Yushi Hayashi, Data Acquisition, Writing-Review & Editing. Leslie W. Liberman Data Acquisition, Writing-Review & Editing. Emily Nguyen, Data Curation, Writing-Review & Editing. Erdem Yildiz, Data Acquisition, Writing-Review & Editing. Jinkyung Kim, Data Acquisition, Writing-Review & Editing. Amandine Jarysta, Data Acquisition, Writing-Review & Editing. Justine Renauld, Data Acquisition, Writing-Review & Editing. Ella Wesson, Data Acquisition, Writing-Review & Editing. Haobing Wang, Data Acquisition, Writing-Review & Editing. Punam Thapa, Data Acquisition, Writing-Review & Editing Pierrick Bordiga, Data Acquisition, Writing-Review & Editing. Noah McMurtry, Data Acquisition, Writing-Review & Editing. Juan Llamas, Data Acquisition, Writing-Review & Editing. Siân R. Kitcher, Data Acquisition, Writing-Review & Editing. Ana I. López-Porras, Data Acquisition, Writing-Review & Editing. Runjia Cui, Data Acquisition, Writing-Review & Editing. Ghazaleh Behnammanesh, Data Acquisition, Writing-Review & Editing. Jonathan Bird, Data Acquisition, Writing-Review & Editing. Angela Ballesteros, Data Acquisition, Writing-Review & Editing. A. Catalina Vélez-Ortega, Data Acquisition, Writing-Review & Editing. Albert SB Edge, Data Acquisition, Writing-Review & Editing. Michael R. Deans, Data Acquisition, Writing-Review & Editing. Ksenia Gnedeva, Data Acquisition, Writing-Review & Editing. Brikha R. Shrestha, Data Acquisition, Writing-Review & Editing. Uri Manor, Data Acquisition, Writing-Review & Editing. Bo Zhao, Data Acquisition, Writing-Review & Editing. Anthony J. Ricci, Data Acquisition, Writing-Review & Editing. Basile Tarchini, Data Acquisition, Writing-Review & Editing. Martín L. Basch, Data Acquisition, Writing-Review & Editing. Ruben Stepanyan, Data Acquisition, Writing-Review & Editing. Lukas D. Landegger, Data Acquisition, Writing-Review & Editing. Mark A. Rutherford, Data Acquisition, Writing-Review & Editing. M. Charles Liberman, Data Acquisition, Writing-Review & Editing. Bradley J. Walters, Data Acquisition, Writing-Review & Editing. Corné J. Kros, Data Acquisition, Writing-Review & Editing. Guy P. Richardson, Data Acquisition, Writing-Review & Editing. Lisa L.

Cunningham, Data Acquisition, Writing-Review & Editing. Artur A. Indzhykulian, Conceptualization, Project administration, Supervision, Writing – Review & Editing, Resources, Funding acquisition.

## Competing interests
The authors declare no competing interests.

## Additional information
**Correspondence** and requests for materials should be addressed to A.A.I.

**Reprints and permissions information** is available at www.nature.com/reprints.

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



**Christopher J. Buswinka[1,2,3], David B. Rosenberg[1,2,4], Rubina G. Simikyan[1], Richard T. Osgood[1,2,5], Katharine Fernandez[6], Hidetomi Nitta[1], Yushi Hayashi[1,2], Leslie W. Liberman[1,2], Emily Nguyen[1], Erdem Yildiz[7], Jinkyung Kim[8,9,10], Amandine Jarysta[11], Justine Renauld[12], Ella Wesson[1], Haobing Wang[1,2], Punam Thapa[13], Pierrick Bordiga[1,2], Noah McMurtry[14], Juan Llamas[15,16], Siân R. Kitcher[5], Ana I. López-Porras[17], Runjia Cui[18], Ghazaleh Behnammanesh[19], Jonathan E. Bird[19], Angela Ballesteros[18], A. Catalina Vélez-Ortega[17], Albert S. B. Edge[1,2], Michael R. Deans[20,21], Ksenia Gnedeva[15,16], Brikha R. Shrestha[1,2], Uri Manor[4,22], Bo Zhao[14], Anthony J. Ricci[9,23], Basile Tarchini[11,24,25], Martín L. Basch[12], Ruben Stepanyan[12,26], Lukas D. Landegger[7,10], Mark A. Rutherford[27], M. Charles Liberman[1,2,3], Bradley J. Walters[13], Corné J. Kros[5], Guy P. Richardson[5], Lisa L. Cunningham[5] & Artur A. Indzhykulian[1,2,3] ✉**

[1]Eaton Peabody Laboratories, Mass Eye and Ear, Boston, MA, 02114, USA. [2]Department of Otolaryngology, Head and Neck Surgery, Harvard Medical School, Boston, MA, 02114, USA. [3]Speech and Hearing Biosciences and Technology graduate program, Harvard University, Cambridge, MA, 02138, USA. [4]Department of Cell and Developmental Biology, University of California San Diego, La Jolla, CA, 92093, USA. [5]Sussex Neuroscience, School of Life Sciences, University of Sussex, Brighton, UK. [6]Section on Sensory Cell Biology, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, 20814, USA. [7]Department of Otolaryngology, Head and Neck Surgery, Vienna General Hospital and Medical University of Vienna, 1090, Vienna, Austria. [8]Department of Otolaryngology, Washington University School of Medicine, St. Louis, MO, 63110, USA. [9]Department of Neuroscience, Washington University School of Medicine, St. Louis, MO, 63110, USA. [10]Department of Otolaryngology, Stanford University School of Medicine, Stanford, CA, 94305, USA. [11]The Jackson Laboratory, Bar Harbor, ME, 04609, USA. [12]Department of Otolaryngology-Head and Neck Surgery, Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA. [13]The University of Mississippi Medical Center, Department of Otolaryngology – Head and Neck Surgery, Jackson, MS, 39216, USA. [14]Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN, 46202, USA. [15]Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, 90033, USA. [16]Tina and Rick Caruso Department of Otolaryngology-Head and Neck Surgery, University of Southern California, Los Angeles, CA, 90033, USA. [17]Department of Physiology, University of Kentucky, Lexington, KY, 40536, USA. [18]Section on Sensory Physiology and Biophysics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, 20814, USA. [19]Department of Pharmacology and Therapeutics, University of Florida, Gainesville, FL, 32610, USA. [20]Department of Neurobiology, Spencer Fox Eccles School of Medicine at the University of Utah, Salt Lake City, UT, 84112, USA. [21]Department of Otolaryngology – Head & Neck Surgery, Spencer Fox Eccles School of Medicine at the University of Utah, Salt Lake City, UT, 84132, USA. [22]Waitt Advanced Biophotonics Center, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA, 92037, USA. [23]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, 94305, USA. [24]Tufts University School of Medicine, Boston, 02111, MA, USA. [25]Graduate School of Biomedical Science and Engineering (GSBSE), University of Maine, Orono, ME, 04469, USA. [26]Department of Neurosciences, Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA. [27]Department of Otolaryngology, Washington University, 660 S. Euclid Avenue, Campus Box 8115, St. Louis, MO, 63110, USA. ✉e-mail: artur_indzhykulian@hms.harvard.edu