The Hair Cell Analysis Toolbox: A machine learning-based whole cochlea analysis pipeline.

Christopher J. Buswinka, Richard T. Osgood, Rubina G. Simikyan, David B. Rosenberg, Artur A. Indzhykulian

Mass Eye and Ear, Harvard Medical School.

* Corresponding author: Artur Indzhykulian, inartur@hms.harvard.edu

Abstract. Our sense of hearing is mediated by sensory hair cells, precisely arranged and highly specialized cells subdivided into two subtypes: outer hair cells (OHCs) which amplify sound-induced mechanical vibration, and inner hair cells (IHCs) which convert vibrations into electrical signals for interpretation by the brain. One row of IHCs and three rows of OHCs are arranged tonotopically; cells at a particular location respond best to a specific frequency which decreases from base to apex of the cochlea. Loss of hair cells at a specific place affects hearing performance at the corresponding tonotopic frequency. To better understand the underlying cause of hearing loss in patients (or experimental animals) a plot of hair cell survival along the cochlear frequency map, known as a cochleogram, can be generated post-mortem, involving manually counting thousands of cells. Currently, there are no widely applicable tools for fast, unsupervised, unbiased, and comprehensive image analysis of auditory hair cells that work well either with imaging datasets containing an entire cochlea or smaller sampled regions. Current microscopy tools allow for imaging of auditory hair cells along the full length of the cochlea, often yielding more data than feasible to manually analyze. Here, we present a machine learning-based hair cell analysis toolbox for the comprehensive analysis of whole cochleae (or smaller regions of interest). The Hair Cell Analysis Toolbox (HCAT) is a software that automates common image analysis tasks such as counting hair cells, classifying them by subtype (IHCs vs OHCs), determining their best frequency based on their location along the cochlea, and generating cochleograms. These automated tools remove a considerable barrier in cochlear image analysis, allowing for faster, unbiased, and more comprehensive data analysis practices. Furthermore, HCAT can serve as a template for deep-learning-based detection tasks in other types of biological tissue: with some training data, HCAT's core codebase can be trained to develop a custom deep learning detection model for any object on an image.

Keywords: cochlea, hair cell, automated analysis, machine-learning, cochleogram.

Introduction

The cochlea is the organ in the inner ear responsible for the detection of sound. It is tonotopically organized in an ascending spiral, with mechanosensitive sensory cells responding to high frequency sounds at its base, and low frequency sounds at the apex. These mechanically sensitive cells of the cochlea, known as hair cells, are classified into two functional subtypes: outer hair cells (OHC) which amplify sound vibrations, and inner hair cells (IHC) which convert these vibrations into neural signals[1]. Each hair cell carries a bundle of actin-rich microvillus-like protrusions called stereocilia. Hair cells are regularly organized into one row of IHCs and three (rarely four) rows of OHCs within a sensory organ known as the Organ of Corti[2]. The OHC stereocilia bundles are arranged in a characteristic V-shape and are composed of thinner stereocilia as compared to those of IHCs. Hair cells are essential for hearing, and deafness phenotypes are often characterized by their histopathology using high-magnification microscopy. The cochlea contains thousands of hair cells, organized over a large spatial area along the length of the Organ of Corti. During histological analysis, each of these thousands of cells represents a datum which must be parsed from the image by hand ad nauseam. To accommodate for manual analysis, it is common to disregard all but a small subset of cells, sampling large datasets in representative tonotopic locations (often referred to as base, middle and apex of the cochlea). To our knowledge, there are two existing automated hair cell counting algorithms to date, both of which have been developed for specific use cases, largely limiting their application for the wider hearing research community. One such algorithm by Urata *et al*[3]. relies on the homogeneity of structure in the organ of Corti and fails when irregularities, such as four rows of outer hair cells, are present. Another one, developed by Cortada *et al*[4] does not differentiate between inner and outer hair cells. Thus, each were limited in their application, likely impeding their widespread use[3,4]. The slow speed and tedium of manual analysis poses a significant barrier when faced with large datasets, be that analyzing whole cochlea instead of sampling three regions, or those generated through studies involving high-throughput screening[5,6]. Furthermore, manual analyses can be fraught with user error, biases, sample-to-sample inconsistencies, and variability between individuals performing the analysis. These challenges highlight a need for unbiased, automated image analysis on a single-cell level across the entire frequency spectrum of hearing.

Over the past decade, considerable advancements have been made in deep learning approaches for object detection[7]. The predominant approach is Faster R-CNN[8], a deep learning algorithm which quickly recognizes the location and position of objects in an image. While originally designed for use with images collected by conventional means (camera), there has been success in applying the same architecture to biomedical image analysis tasks[9-11]. This algorithm can be adapted and trained to perform such tasks orders of magnitude faster than manual analysis. We have created a machine-learning-based analysis software which quickly and automatically *detects* each hair cell, determines its *type* (IHC vs OHC), and estimates cell's *best frequency* based on its location along the cochlear coil. Here, we present a suite of tools for cochlear hair cell image analysis, the Hair Cell Analysis Toolbox (HCAT), a consolidated software that enables fully unsupervised hair cell detection and cochleogram generation.

Results

Analysis Pipeline: HCAT combines a deep learning algorithm, which has been trained to detect and classify cochlear hair cells, with a novel procedure for cell frequency estimation to extract information from cochlear imaging datasets quickly and in a fully automated fashion. An overview of the analysis pipeline is shown in **Figure 1.** The model accepts common image formats (tif, png, jpeg), in which the order of the fluorescence channels within the images, or their assigned color, does not affect the outcome. Multi-page tif images are automatically converted to a 2D maximum intensity projection. When working with large confocal micrographs, HCAT analyzes small crops of the image and subsequently merges the results to form a contiguous detection dataset. These cropped regions are set to have 10% overlap along all edges, ensuring that each cell is fully represented at least once. Regions which do not contain any fluorescence above a certain threshold may be optionally skipped, increasing speed of large image analysis while limiting false positive errors. When the entire cochlea is contained as a contiguous piece (**Figure 1a**), which is common for neonatal cochlear histology, HCAT will estimate the cochlear path and each cell will be assigned a best frequency. Following cell detection and best frequency estimation, HCAT performs two post-processing steps to refine the output and improve overall accuracy. First, cells detected multiple times are identified and removed based on a user-defined bounding box overlap threshold, set to 30% by default. The second step, optional and only applicable for whole cochlear coil analysis, removes cells too far from the estimated cochlear path, reducing false-positive detections in datasets with sub-optimal anti-MYO7A labeling outcomes, such as high background fluorescence levels or instances of non-specific labeling away from the Organ of Corti. As outlined below, for each detection analysis HCAT outputs diagnostic images with overlaid cell-specific data, in addition to an associated CSV data table, enabling further data analysis or downstream postprocessing, and, when applicable, automatically generates cochleograms.

HCAT is computationally efficient and can execute detection analysis on a whole cochlea on a timescale vastly faster than manual analysis, regularly completing in under 90 seconds when utilizing GPU acceleration on affordable computational hardware. HCAT is available in two user interfaces: 1) a command line interface which offers full functionality, including cell frequency estimation and batch processing of multiple images or image stacks across multiple folders and 2) a graphical user interface (GUI), which is user-friendly and is optimized for analysis of individual or multiple images contained within a single folder. The GUI interface is unable to infer cell's best frequency and is suitable for analysis of small regions of cochlea.

Detection and Classification: To perform cell detection, we leverage the Faster R-CNN[8] deep learning algorithm with a ConvNext[12] backbone trained on a varied dataset of cochlear hair cells from multiple species, at different ages, and from different experimental conditions (**Table 1, Figure 2**). Most of the hair cells used to train the detection model were stained with two markers: (1) anti-MYO7A, a hair cell specific cell body marker and (2) the actin label, phalloidin, to visualize the stereocilia bundle. Bounding boxes for each cell along with class identification labels were manually generated to serve as the ground truth reference by which we trained the detection model (**Figure 2**). Boxes were centered around stereocilia bundles and included the hair cell cuticular plate as these were determined the most robust features per cell in a maximum intensity projection image. The trained Faster R-CNN model predicts three features for each detected cell: a bounding box, a classification label (IHC or OHC), and a confidence score (**Figure 3**).

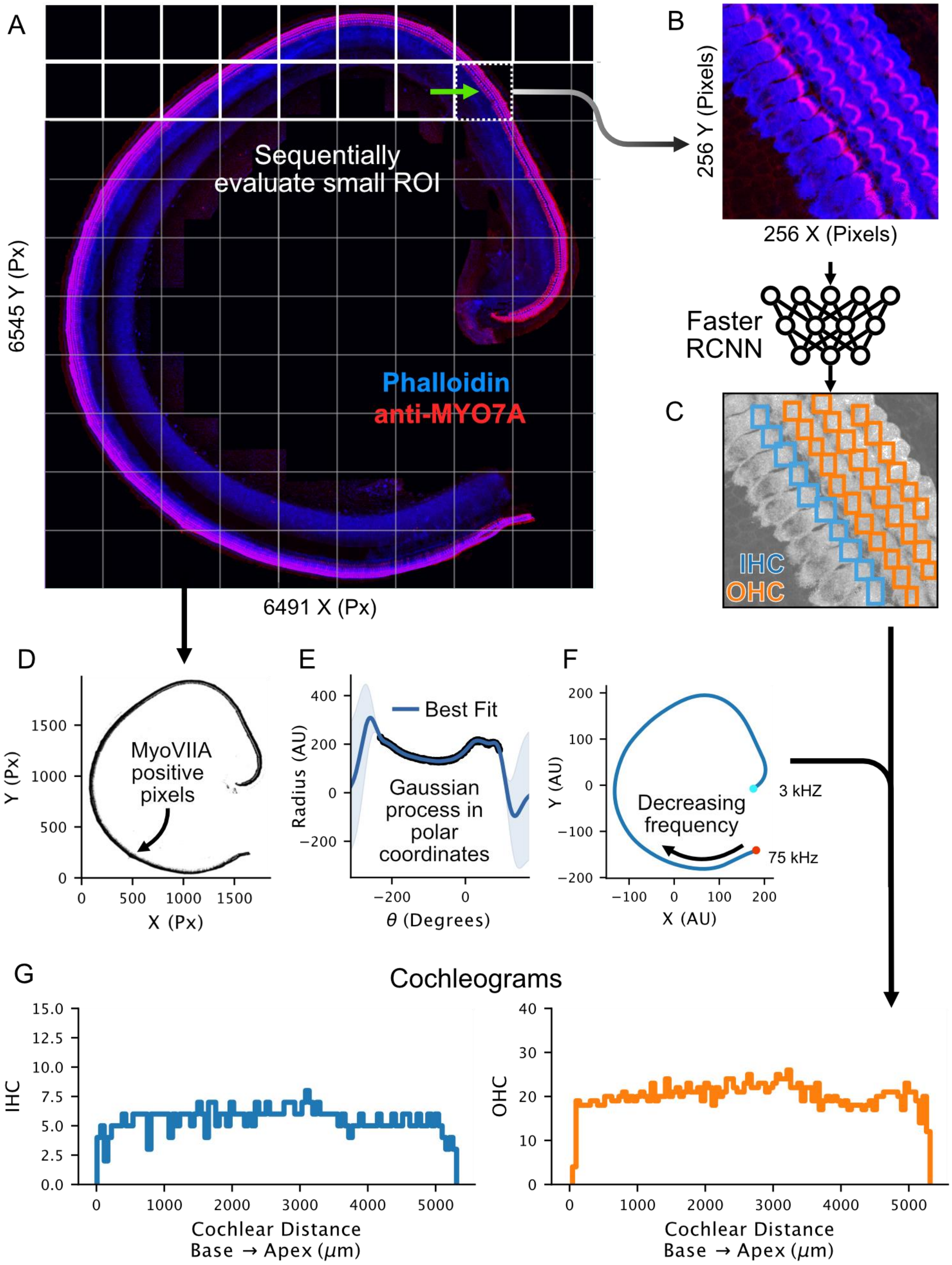


Figure 1. **HCAT Analysis Pipeline:** A whole cochlea imaged at high magnification (288 nm/px resolution) (**A**) is broken into smaller 256x256 px regions and sequentially evaluated by a deep learning detection and classification algorithm (**B**) to predict the probable locations of inner and outer hair cells (**C**). The entire cochlea is then used to infer each cell's best frequency along the cochlear coil. First, all supra-threshold anti-MYO7A-positive pixels are converted to polar coordinates (**D**) and fit by the Gaussian process nonlinear curve fitting algorithm (**E**). The resulting curve is converted back to cartesian coordinates and the resulting line is converted to frequency by the Greenwood function; the apical end of the cochlea (*teal circle*) is inferred by the region of greatest curl (**F**), and the opposite end of the cochlea is assigned as the basal end (*red circle*). Cells are then assigned a best frequency based on their position along the predicted curve, and cochleograms (**G**) are generated in a fully automated way for each cell type (IHCs and OHCs), with a bin size by default set to 1% of the total cochlear length.

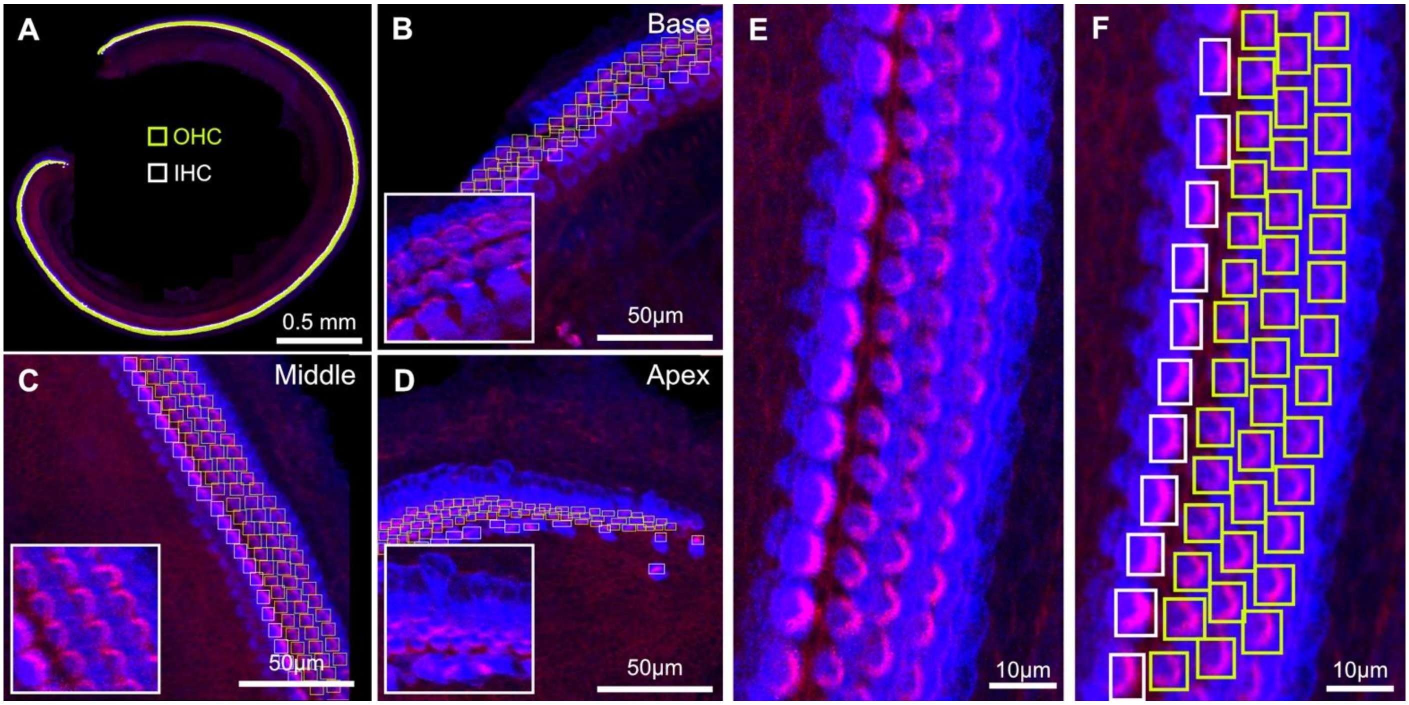


Figure 2. **HCAT detection algorithm training data.** Hair cells in whole cochlea stained against MYO7A (*blue*) and phalloidin (*magenta*) were manually annotated (**A-D**) and used as training data for the Faster R-CNN deep learning algorithm. Hair cells vary in appearance based on tonotopy, with representative regions of the base (**B**), middle (**C**), and apex (**D**) shown here. Since the boundaries between hair cell cytosol (*blue*) overlap in maximum intensity projection images (**E**), the bounding boxes for each cell were annotated around the stereocilia bundle and cuticular plate of each hair cell (**F**).

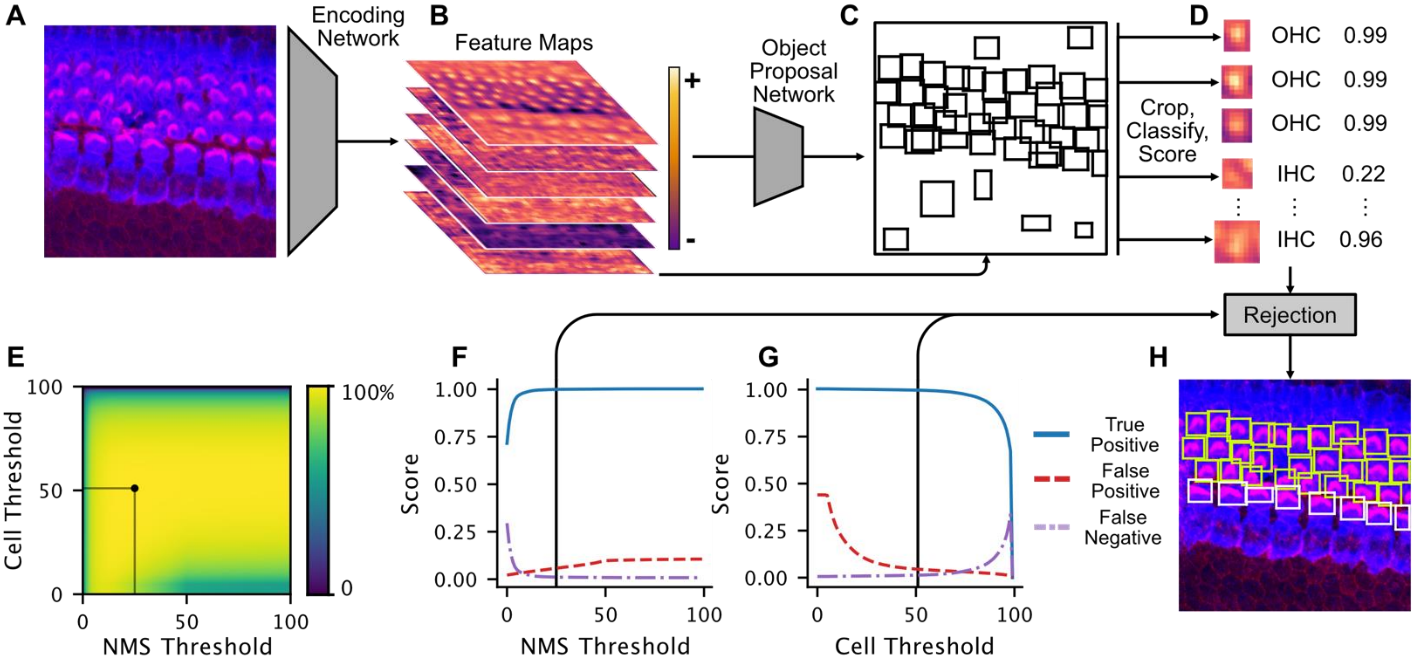


Figure 3. Overview of Faster R-CNN image detection backend. (**A**), Input micrographs are encoded into high-level representations (schematized in **B**) by a trained encoding convolutional neural network. These high-level representations are next passed to a region proposal network which predicts bounding boxes of objects (**C**). Based on the predicted object proposals, encoded crops are classified into OHC and IHC classes, and assigned a confidence score (**D**). Next, a rejection step thresholds the resulting predictions based on confidence scores and the overlap between boxes, via non-maximum suppression (NMS). Default values for user-definable thresholds were determined by the maximum average precision after a grid search of parameter combinations over eight manually annotated cochleae (**E**). The outcome of this grid search can be flattened into accuracy curves for the NMS (**F**) and rejection threshold (**G**) at their respective maxima. Boxes remaining after rejection represent the models' best estimate of each detected object in the image (**H**).

To limit false positive detections, cells predicted by Faster R-CNN can be rejected based on their confidence score, or their overlap with another detection through an algorithm called non-maximum suppression (NMS). To find optimal values for the confidence and overlap thresholds, we performed a grid search by which we assessed model performance at each combination of values and selected values leading to most accurate model performance (**Figure 3E, F, G,** and **Supplemental Figure S1**).

Table 1. Summary of training data.

Laboratory	Number of images	OHC	IHC	Animal	Microscope	Treatment	Labeled Protein	
Artur Indzhykulian, PhD	45	12959	3706	Mouse	Confocal	None	MYO7A	Actin
Lisa Cunningham, PhD	77	3424	1290	Mouse	Confocal	Platinum Compounds	MYO7A	Actin
Albert Edge, PhD	2	125	42	Mouse	Confocal	None	MYO7A	Actin
M. Charles Liberman, PhD	29	894	290	Human	Confocal	None	MYO7A	ESPN
Guy Richardson, PhD and Corne Kros, PhD	26	1226	690	Mouse	Epifluorescence	Aminoglycosides	MYO7A	Actin
Mark Rutherford, PhD	5	120	65	Mouse	Confocal	None	MYO7A	Actin
Anthony Ricci, PhD	2	120	43	Mouse	Confocal	None	MYO7A	Actin
Basile Tarchini, PhD	8	292	97	Mouse	Confocal	None	MYO7A	Actin
Bradley Walters, PhD	6	904	238	Guinea Pig	Confocal	None	MYO7A	Actin
Total	200	20064	6461					

The trained Faster R-CNN detection algorithm performs best on maximum intensity projections of 3D confocal z-stacks of hair cells labelled with a cell body stain (such as anti-MYO7A) and a hair bundle stain (such as phalloidin), imaged at a X-Y resolution of ~290 nm/px (**Figure 4D, E**). However, the model can perform well with combinations of other markers, including antibody labeling against ESPN, Calbindin, Calcineurin, p-AMPK**α**, as well as following FM1-43 dye loading. HCAT can accurately detect cells in healthy and pathologic cochlear samples, collected within a range of imaging modalities, resolutions, and signal-to-noise ratios. While the pixel resolution requirements for the imaging data are not very demanding, imaging artifacts and low fluorescence signal intensity can limit detection accuracy. Although there is one row of IHCs and three rows of OHCs in most cochlear samples, there are rare instances where two rows of IHCs or four rows OHCs can be seen in normal cochlear samples, the algorithm is robust and largely accurate in such instances (**Figure 4D**).

Cochlear path determination: For images containing an entire contiguous cochlear coil, HCAT can additionally predict cell's best frequency via automated cochlear path determination. To do this, HCAT fits a Gaussian process nonlinear regression[13] through the ribbon of anti-MYO7A-positive pixels, effectively treating each hair cell as a point in cartesian space. A line of best fit can be predicted through each hair cell and in doing so approximate the curvature of the cochlea. The length of this curve is then used as an approximation for the length of the cochlear coil. For example, a cell that is 20% along the length of this curve could be interpreted as one positioned at 20% along the length of the cochlea, assuming the entire cochlear coil was imaged.

To optimally perform the initial regression, individual cell detections are rasterized and then downsampled by a factor of ten using local averaging (increasing the execution speed of this step), then converted to a binary image. Next, a binary hole closing operation is used to close any gaps, and subsequent binary erosion is used to reduce the effect of nonspecific staining. Each positive binary pixel of the resulting two-dimensional image is then treated as an X/Y coordinate which may be regressed against (**Figure 1D**). The resulting image is unlikely to form a mathematical function in cartesian space however, as the cochlea may curve over itself such that for a single location on the X axis, there may be multiple clusters of cells at different Y values. To rectify this overlap, the data points are converted from cartesian to polar coordinates by shifting the points and centering the cochlear spiral around the origin, then converting each X/Y coordinate to a corresponding angle/radius coordinate. As the cochlea is not a closed loop, the resulting curve will have a gap, which is then detected by the algorithm, shifting these points by one period, and creating a continuous function. A Gaussian process[13], a generalized nonlinear function, is then fit to the polar coordinates and a line of best fit is predicted. This line is then converted back to cartesian coordinates and scaled up to correct for the earlier down-sampling (**Figure 1E**).

The apex of the cochlea is then inferred by comparing the curvature at each end of the line of best fit based on the observation that the apex has a tighter curl when mounted on a slide. The resulting curve closely tracks the hair cells on the image. Next, the curve's length is measured, and each detected cell is then mapped to it as a function of the total cochlear length (%). Each cell's best frequency is calculated using the Greenwood function, a species-specific method of determining cell's best frequency from its cochlear position[14] (**Figure 1F**). Upon completion of this analysis, the automated frequency assignment tool generates two cochleograms, one for IHCs and one for OHCs (**Figure 1G**).

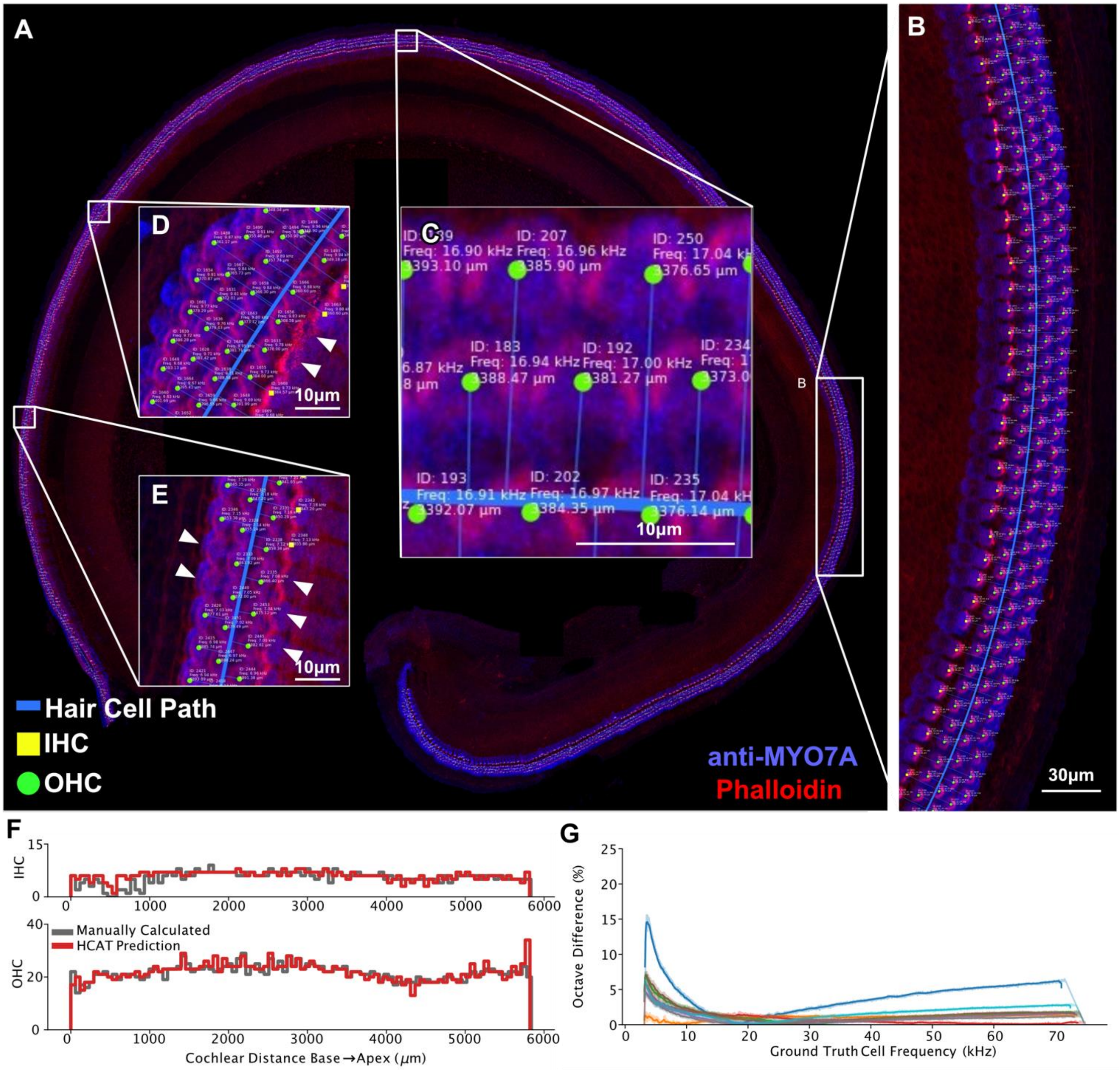


Figure 4. Validation output of the hair cell detection analysis. A validation output image is generated for each detection analysis performed by the software. An image is automatically generated by the software similar to the one shown here for a dataset that includes an entire cochlea (**A**), with the vast majority of cells accurately detected (**B**). For each image, the model embeds information on cell's ID, its location along the cochlear coil (distance in μm from the apex), it's best frequency, cell classification (IHC as yellow squares, OHC as green circles) and the line that represents tool's cochlear path estimation (**C,** *blue line*). The very few examples of poor performance are highlighted in **D** and **E** (arrowheads point to 3 missed IHCs and two OHCs). A set of cochleograms reporting cell counts per every 1% of total cochlear length, generated with manual cell counts and frequency assignment (*grey*) closely agrees with a HCAT-predicted cochleogram (*red*) generated in a fully automated fashion (**F**). To assess the accuracy of tool's best frequency assignment, the magnitude difference between every cell's best frequency calculated manually, and automatically, with respect to frequency for eight different cochleae is at maximum 15% of an octave across all frequencies (**G**). Each color represents one cochlea.

To validate this method of best frequency assignment, we compared it to the existing standard in the field – manual frequency estimation. We manually mapped the cochlear length to cochlear frequency using a widely used *imageJ* plugin, developed by the Histology Core at the Eaton-Peabody Laboratories (Mass Eye and Ear) and compared them to the results predicted by our automatic tool (**Figure 4G** and **Supplemental Figure S1**). Over eight manually analyzed cochleae, the *maximum* cell frequency error of automated, relative to a manually, mapped best frequency was under 10% of an octave, with the discrepancy between the two methods less than 5% for most cells (60% of a semitone). In one cochlea, the overall cochlear path was predicted to be shorter than manually assigned, due to the threshold settings of the MYO7A channel, causing an error at very low and very high frequencies (**Figure 4G,** *dark blue*). While this error was less than 0.15% of an octave, it is an outlier in the dataset. It is recommended, when using this tool, to evaluate the automated cochlear path estimation, and if poor, perform manual curve annotation to facilitate best frequency assignment. If required, the user is also able to switch the designation of automatically detected points representing the apical and basal ends of the cochlear coil (**Figure 1F**, *red* and *cyan* circles).

<u>*Performance:*</u> Overall, cochleograms generated with HCAT track remarkably well to those generated manually (**Figure 4F**). Comparing HCAT to manually annotated cochlear coils (not used to train the model), we report a 98.6±0.005% true positive accuracy for cell identification and a <0.01% classification error (8 cochlear coils, 4428 IHCs and 15754 OHCs; **Supplemental Figure S1**). We found no bias in accuracy with respect to estimated best frequency. To assess HCAT performance on a diverse set of cochlear micrographs, we sampled 88 images from 15 publications[15-29] that represent a wide variety of experimental conditions, including ototoxic treatment using aminoglycosides, genetic manipulations that could affect the hair cell anatomy, noise exposure, blast trauma and age-related hearing loss (**Table 2**). We performed a manual quantification and automated detection analysis of these images after they were histogram-adjusted and scaled via the HCAT GUI for optimal accuracy. HCAT achieved an overall OHC detection accuracy of 98.6±0.5% and an IHC detection accuracy of 96.9±2.8% for 3545 OHCs and 1110 IHCs, with mean error of 0.34 OHC and 0.32 IHC per image. Of the 88 images we used for this validation, no errors were detected on 62 of them, and HCAT was equally accurate in images of low and high absolute cell count (**Figure 5).**

Table 2. Summary of micrographs sampled from existing publications to test HCAT performance.

Lab	Number of images	OHC	IHC	Animal	Microscopy	Treatment	Age	Labeled Protein	
Beurg et al., 2019	2	39	17	Mouse	Confocal	*Tmc1*$^{\mathrm{p.D569N}}$ mouse	Neonatal	Calbindin	Actin
Fang et al., 2019	1	42	14	Mouse	Confocal	WT mouse	Adult	MYO7A	Actin
Fu et al., 2022	6	175	69	Mouse	Confocal	*Klc2*$^{-/-}$ mouse	Adult	MYO7A	Actin
Gyorgy et al., 2019	11	330	113	Mouse	Confocal	*Tmc1*^{Bth} mutant	Adult	MYO7A	Actin
He et al., 2021	7	304	69	Mouse	Confocal	Noise trauma	Adult	Calcineurin	Actin
Hill et al., 2016	5	171	0	Mouse	Confocal	Noise trauma	Adult	p-AMPK**α**	Actin
Kim et al., 2018	3	102	33	Mouse	Confocal	Blast trauma	Adult	MYO7A	Actin
Lee et al., 2017	2	24	7	Mouse	Confocal	WT mouse	Neonatal	MYO7A	Actin
Li et al., 2020	2	70	21	Mouse	Confocal	*Myo7a-ΔC* mouse	Adult	MYO7A	Actin
Mao et al., 2021	9	916	311	Mouse	Confocal	Blast trauma	Adult	MYO7A	Actin
Sang et al., 2015	6	193	65	Mouse	Confocal	*Idlr1*$^{-/-}$ mouse	Adult	MYO7A	Actin
Sethna et al., 2021	7	274	104	Mouse	Confocal	*Pcdh15*^{R250X} mouse	Adult	MYO7A	Actin
Wang et al., 2011	2	90	26	Mouse	Confocal	*SCX*$^{-/-}$ mouse	Adult	MYO7A	Actin
Yousaf et al., 2015	24	760	244	Mouse	Confocal	*Map3k1*^{tm1Yxia}	Adult	MYO7A	Actin
Zhao et al., 2021	1	55	17	Mouse	Confocal	*Clu*$^{-/-}$ mouse	Adult	MYO7A	Actin
Total	88	3545	1110						

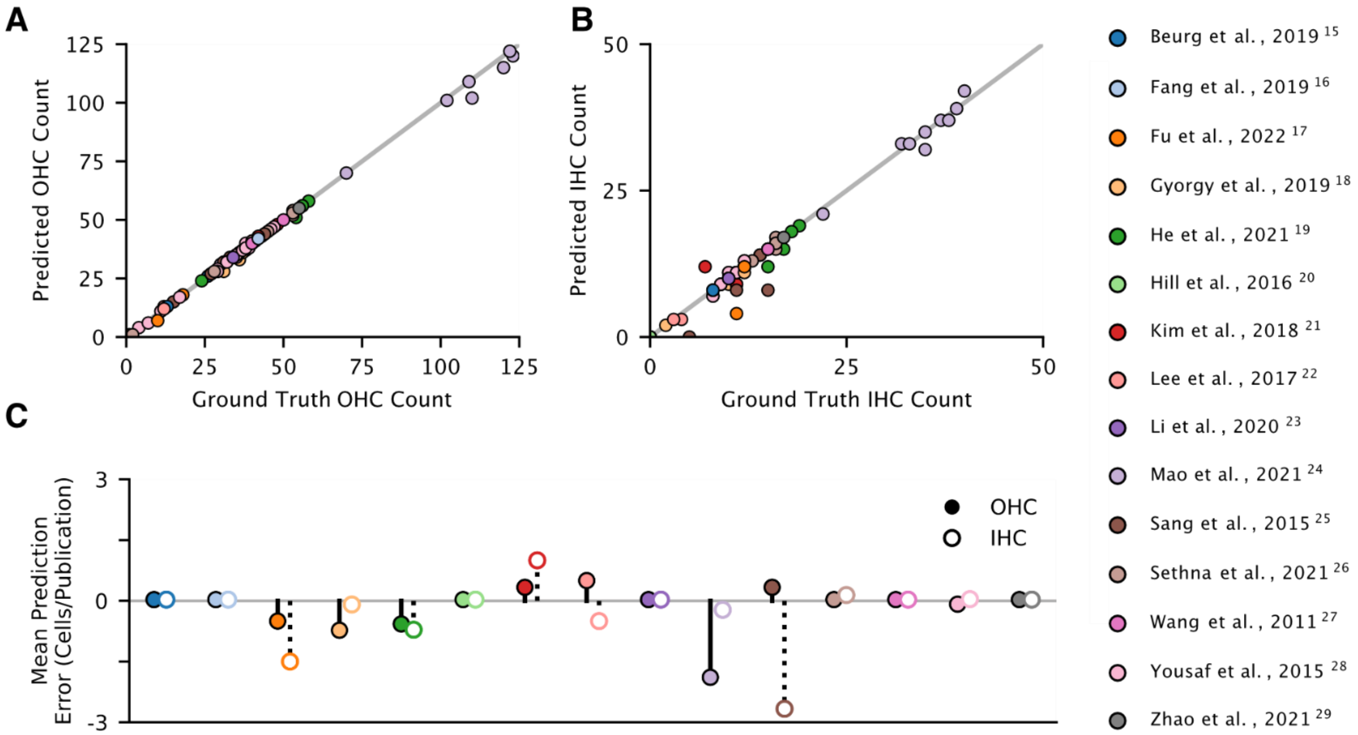


Figure 5. HCAT detection performance on published images of cochlear hair cells. HCAT detection performance was assessed by running a cell detection analysis in the GUI on 88 confocal images of cochlear hair cells sampled from published figures across 15 different original studies[15-29]. None of the images from this analysis were used to train the model. Each image was adjusted within the GUI for optimal detection performance. Cells in each image were also manually counted (presented as ground truth values) and results compared to HCAT's automated detection. The resulting population distributions of hair cells are compared for OHCs (**A**), and IHCs (**B**). The mean difference in predicted number of IHCs (open circles) and OHCs (filled circles) in each publication is summarized for each cell type: zero indicates an accurate detection, negative values indicate false-negative detections, while positive values indicate false-positive detections (**C**).

Validation on published datasets: We further evaluated HCAT on whole, external datasets (generously provided by the Cunningham[30], Richardson and Kros laboratories[6]) and replicated analyses from their publications. Each dataset presented examples of Organ of Corti epithelia treated with ototoxic compounds resulting in varying degrees of hair cell loss. The two datasets complement each other in several ways, covering most use cases of data analysis needs following ototoxic drug use in the Organ of Corti to assess hair cell survival: *in-vivo* vs. *in-vitro* drug application, confocal fluorescence vs. widefield fluorescence microscopy imaging, early postnatal vs. adult Organ of Corti imaging. HCAT succeeded in quantifying the respective datasets in a fully automated fashion with an accuracy sufficient to replicate the main finding in each study (**Figure 6**). It is worth noting that these datasets were collected without optimization for an automated analysis. Thus, we expect an even higher performance accuracy with an experimental design optimized for HCAT-based automated analysis.

Discussion

Here we present the first fully automated cochlear hair cell analysis pipeline for analyzing multiple micrographs of cochleae, quickly detecting and classifying hair cells. HCAT can analyze whole cochleae or individual regions and can be easily integrated into existing experimental workflows. While there were previous attempts at automating this analysis, each were limited in their use to achieve widespread application[3,4]. HCAT allows for unbiased, automated hair cell analysis with detection accuracy levels approaching that of human experts at a speed so significantly faster that it is desirable even with rare errors. Furthermore, we validate HCAT on data from various laboratories and find it is accurate across different imaging modalities, staining, age, and species.

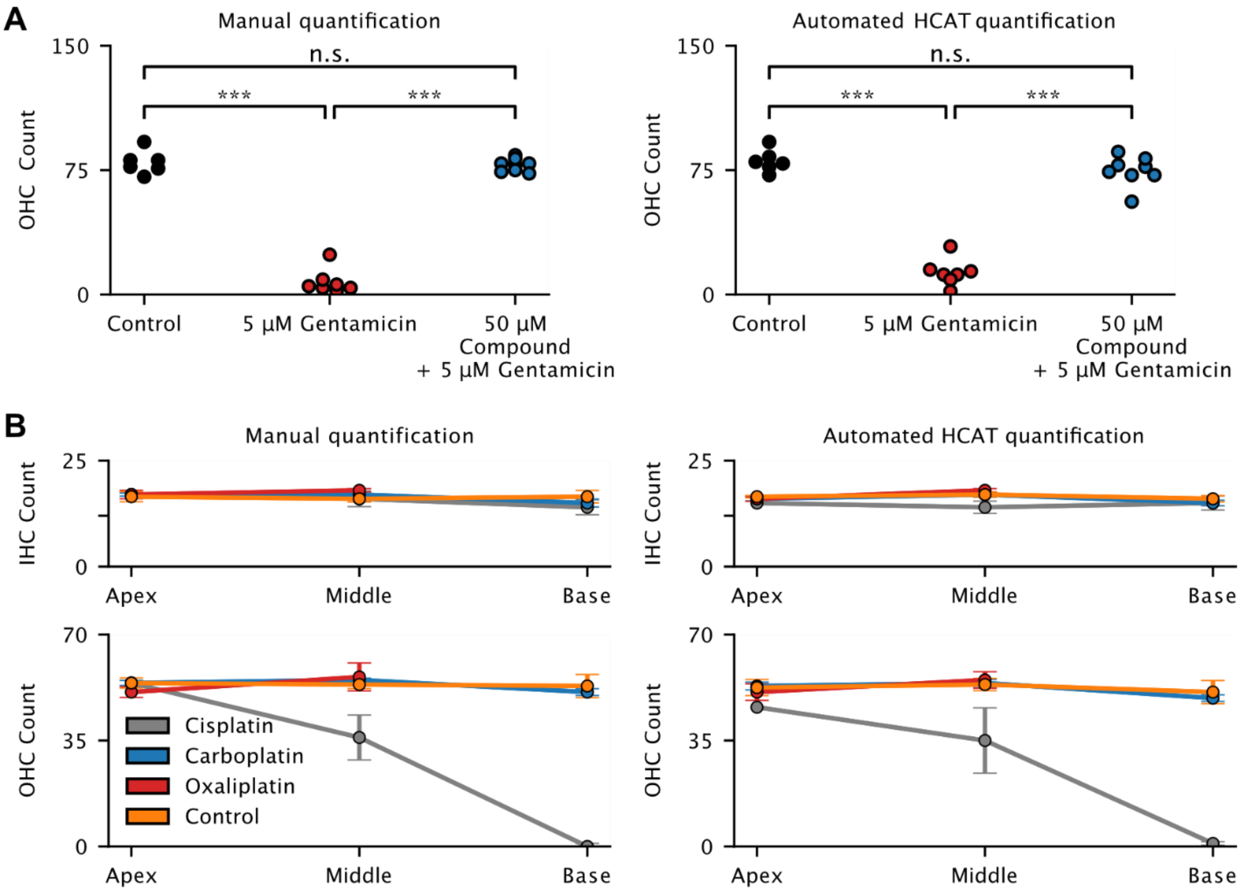


Figure 6. Evaluation of HCAT performance on cochlear datasets to assess ototoxic drug effect. To assess HCAT performance on aberrated cochlear samples, we compared HCAT analysis results to manual quantification on datasets from two different publications focused on assessing hair cell survival following treatment with ototoxic compounds. **(A)** Original imaging data underlying the finding in Figure 2F of Kenyon *et al.,* 2021[6], generously provided by the Richardson and Kros laboratories. Images were collected using epifluorescence microscopy, following a 48-hour incubation in either 0 µM gentamicin (Control), 5 µM gentamicin, or 5 µM gentamicin + 50 µM test compound UoS-7692. Each symbol represents the number of OHCs in a mid-basal region from one early postnatal *in-vitro* cultured cochlea[6]. One-way ANOVA with Tukey's multiple comparison tests. ***, p < 0.001; *ns*, not significant. In some cases, HCAT detections overestimated the total number of surviving hair cells in the gentamycin-treated tissue. However, overall, the software-generated results are in agreement with those of the original study, drawing the same conclusion. **(B)** Original imaging data underlying the finding in Figure 7A-B in Gersten *et al.,* 2020[30] were generously provided by the Cunningham laboratory. In this study, mice were treated by *in-vivo* application of clinically proportional levels of ototoxic compounds, Cisplatin, Carboplatin, Oxaliplatin, and Saline (control), in an intraperitoneally cyclic delivery protocol[30]. Regions of interest were imaged at the base, middle, and apex of each cochlea. HCAT's automated detections were comparable to manual quantification and were sufficient to draw a conclusion that is consistent with the original publication. Upon comparison, HCAT had higher detection accuracy in OHCs, compared to IHCs, likely due to the variability of the MYO7A intensity levels in IHCs across the dataset.

Deep-learning-based detection infers information from the pixels of an image to make decisions about what objects are and where they are located. To this end, the information is devoid of any context. HCAT's deep learning detection model was trained largely using anti-MYO7A and phalloidin labels, however the model can perform on specimens labeled with other markers, as long as they are visually similar to examples in our training data. For example, some of the validation images of cochlear hair cells sampled from published figures contained cell body label other than MYO7A, such as Calbindin[31], Calcineurin[32], and p-AMPKα[33] while in other images phalloidin staining of stereocilia bundle was substituted by anti-espin[34] labeling. Of higher importance is the quality of the imaging data: proper focus adjustment, high signal-to-noise ratio, and adequately adjusted brightness and contrast settings. Furthermore, the quality of the training dataset greatly affects model performance; upon validation, HCAT performed slightly worse when evaluated on community provided datasets due to fewer representative examples within the pool of our training data. We will strive to periodically update our published model when new data arise, further improving performance over time. At present, HCAT has proven to be sufficiently accurate to consistently replicate major findings even with occasional discrepancies to a manual analysis, even when used on datasets that were collected without any optimization for automated analysis. The strength of this software is in automation, allowing for processing thousands of hair cells over the entire cochlear coil without human input.

While the detection model was trained and cochlear path estimation designed specifically for cochlear tissue, HCAT can serve as a template for deep-learning-based detection tasks in other types of biological tissue in the future. While developing HCAT, we employed best practices in model training, data annotation and augmentation. With minimal adjustment and a small amount of training data, one could adapt the core codebase of HCAT to train and apply a custom deep learning detection model for any object in an image.

To our knowledge, this is the first whole cochlear analysis pipeline capable of accurately and quickly detecting and classifying cochlear hair cells. This hair cell analysis toolbox (HCAT) enables expedited cochlear imaging data analysis while maintaining high accuracy. This highly accurate and unsupervised data analysis approach will both facilitate ease of research and improve experimental rigor in the field.

Materials and Methods

Preparation and imaging of in-house training data. Organs of Corti were dissected at P5 in Leibovitz's L-15 culture medium (21083-027, Thermo Fisher Scientific) and fixed in 4% formaldehyde for 1 hour. The samples were permeabilized with 0.2% Triton-X for 30 minutes and blocked with 10% goat serum in calcium-free HBSS for two hours. To visualize the hair cells, samples were labeled with an anti-Myosin 7A antibody (#25-6790 Proteus Biosciences, 1:400) and goat anti-rabbit CF568 (Biotium) secondary antibody. Additionally, samples were labeled with Phalloidin to visualize actin filaments (Biotium CF640R Phalloidin). Samples were then mounted on slides using ProLong® Diamond Antifade Mounting kit (P36965, Thermo Fisher Scientific,) and imaged with a Leica SP8 confocal microscope (Leica Microsystems) using a 63×, 1.3 NA objective. Confocal Z-stacks of 512x512 pixel images with an effective pixel size of 288 nm were collected using the tiling functionality of the Leica LASX acquisition software and maximum intensity projected to form 2D images. All experiments were carried out in compliance with ethical regulations and approved by the Animal Care Committee of Massachusetts Eye and Ear.

Training Data: Despite the National Institutes of Health (NIH) mandate to share NIH-funded data, getting access to imaging data linked to published studies reported by other laboratories remains to be challenging. Varied data are required for the training of generalizable deep learning models. In addition to data collected in our lab, we sourced generous contributions from the larger hearing research community from previously reported [6,30,35-42], and in some cases unpublished, studies. Bounding boxes for hair cells seen in maximum intensity projected z-stacks were manually annotated using the labelImg[43] software and saved as an XML file. For whole cochlear cell annotation, a "human in the loop" approach was taken, first evaluating the deep learning model on the entire cochlea, visually inspecting it, then manually correcting errors. Our dataset contained examples from three different species, multiple ages, microscopy types, and experimental conditions. A summary of our training data is presented in **Table 1**.

Training Procedure: The deep learning architectures were trained with the AdamW[44] optimizer with a learning rate starting at 1e-4 and decaying based on cosine annealing with warm restarts with a period of 10000 epochs. In cases with a small number of training images, deep learning models tend to fail to generalize and instead "memorize" the training data. To avoid this, we made heavy use of image transformations which randomly add variability to the original set of training images and synthetically increase the variety of our training data sets[45] (**Supplemental Figure S2**).

Hyperparameter Optimization: Eight manually annotated cochleae were evaluated with the Faster R-CNN detection algorithm without either rejection method (via detection confidence or non-maximum suppression). A grid search was performed by breaking each threshold value into 100 steps from zero to one, and each combination applied to the resulting cell detections, reducing their number, then calculating a the true positive (TP), true negative (TN), and false positive (FP) rates (**Supplemental Figure S1D-E**). An accuracy metric of the TP minus both TN and FP was calculated and averaged for each cochlea. The combination of values which produce the highest accuracy metric were then chosen as default for the HCAT algorithm.

Computational Environment: HCAT is operating system agnostic, requires at least 8 GB of system memory, and optionally a NVIDIA GPU with at least 8 GB of video memory to optional GPU acceleration. All scripts were run on an analysis computer running Ubuntu 20.04.1 LTS, an open-source Linux distribution from Canonical based on Debian. The workstation was equipped with two Nvidia A6000 graphics cards for a total of 98Gb of video memory. Many scripts were custom written in python 3.9 using open source scientific computation libraries including numpy[46], matplotlib, scikit-learn[47]. All deep learning architectures, training logic, and much of the data transformation pipeline was written in pytorch[48] and making heavy use of the torchvision[48] library.

Acknowledgements. We would like to thank Dr. Marcelo Cicconet (Image and Data Analysis Core at Harvard Medical School) and Haobing Wang, MS (Mass Eye and Ear Light Microscopy Imaging Core Facility) for their assistance in this project. We thank Dr. Lisa Cunningham, Dr. Michael Deans, Dr. Albert Edge, Dr. Katharine Fernandez, Dr. Ksenia Gnedeva, Dr. Yushi Hayashi, Dr. Tejbeer Kaur, Dr. Jinkyung Kim, Prof. Corne Kros, Dr. M. Charles Liberman, Dr. Vijayprakash Manickam, Dr. Anthony Ricci, Prof. Guy Richardson, Dr. Mark Rutherford, Dr. Basile Tarchini, Dr. Bradley Walters, the members of their teams and all other research groups, for providing their datasets to evaluate the HCAT. We thank Hidetomi Nitta, Emily Nyguen, and Ella Wesson for their assistance in generating a portion of training data annotations. We also thank Evan Hale, Corena Loeb for critical reading of the manuscript. This work was supported by NIH R01DC020190 (NIDCD), R01DC017166 (NIDCD) and R01DC017166-04S1 "Administrative Supplement to Support Collaborations to Improve the AI/ML-Readiness of NIH-Supported Data" (Office of the Director, NIH) to A.A.I.

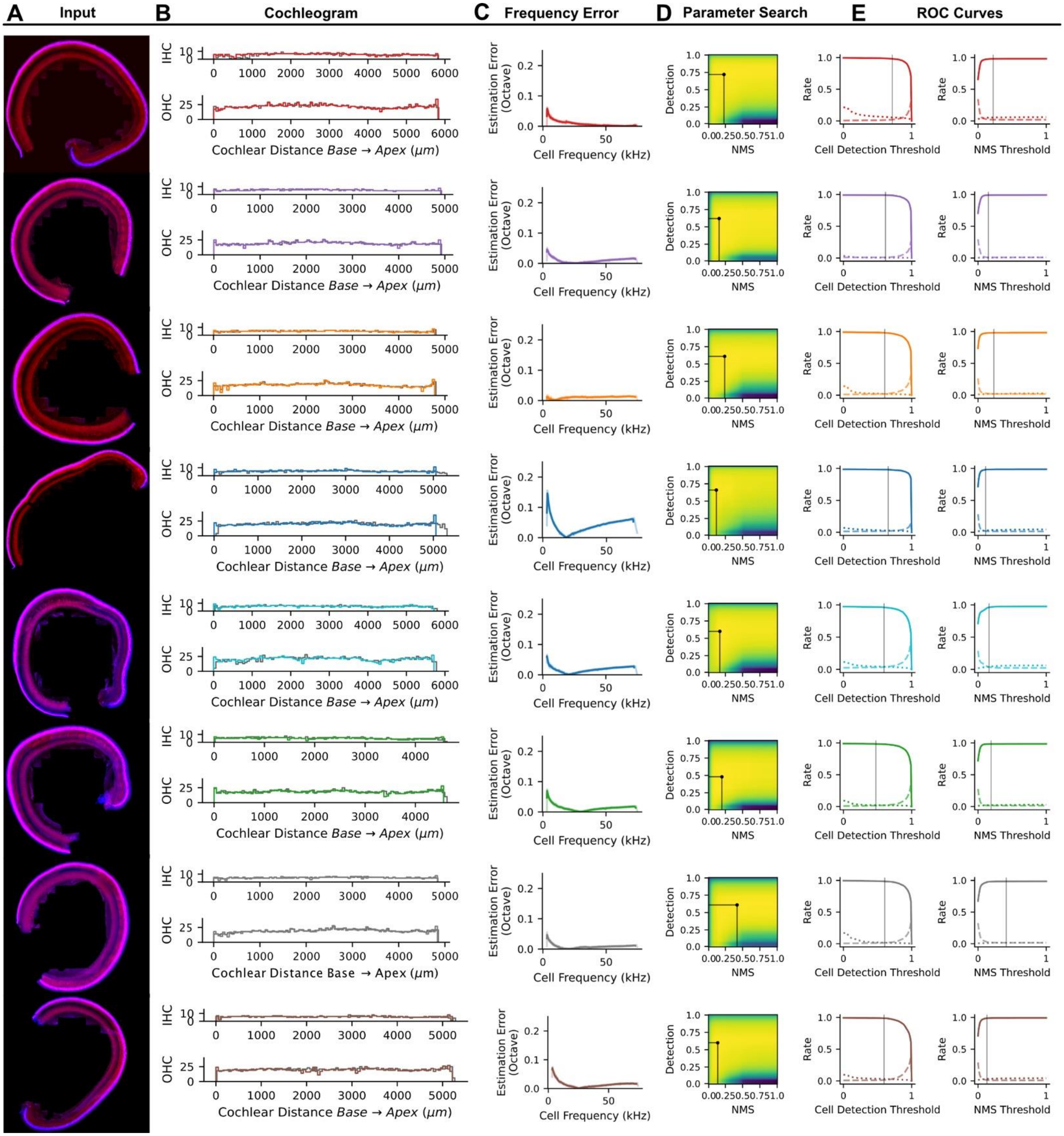
Author contributions.

C.J.B. Conceptualization, Methodology development, Investigation, Data Curation, Software development, Formal analysis, Visualization, Validation, Writing - Original Draft, Resources, Supervision, assistance with Funding acquisition;
R.T.O. Data Curation, Formal analysis, Validation and Visualization (Figure 6A), Writing - Review & Editing;
R.G.S. Data Curation, Writing - Review & Editing;
D.B.R. Investigation/imaging of large portion of in-house training data, Writing - Review & Editing;.
A.A.I. Conceptualization, Methodology development, Visualization, Validation, Writing - Original Draft, Supervision, Project administration, Resources, Funding acquisition. All authors contributed to the final version of the manuscript.

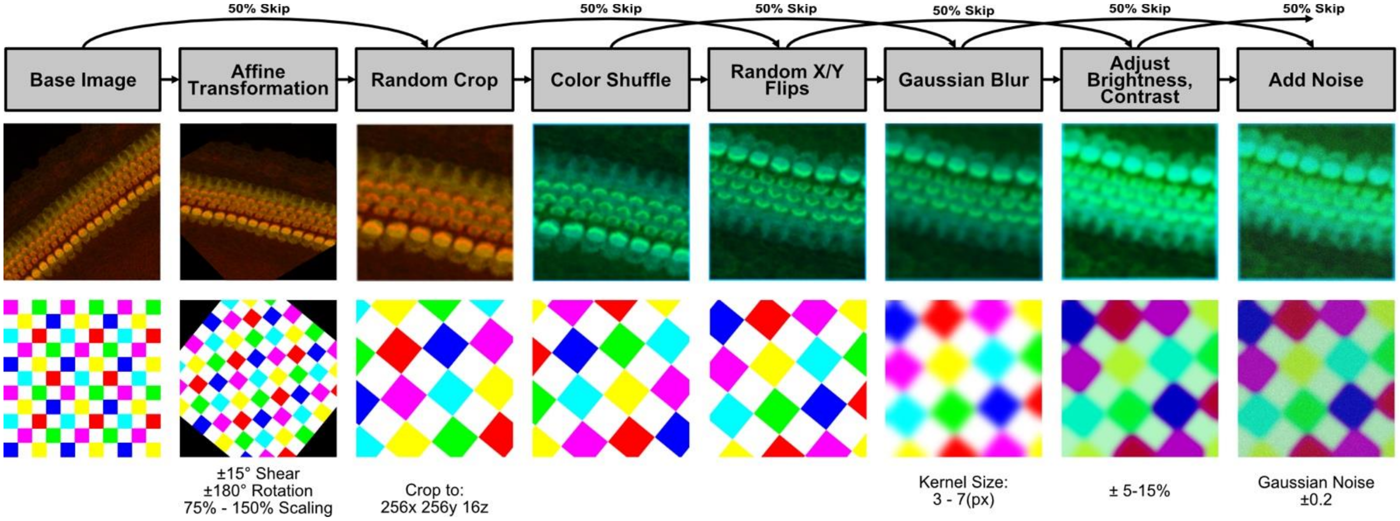
Code availability.

All code has been hosted on github and is available for download at https://github.com/indzhykulianlab/hcat along with accompanying documentation at hcat.readthedocs.io. The EPL cochlea frequency ImageJ plugin is available for download at: https://www.masseyeandear.org/research/otolaryngology/eaton-peabody-laboratories/histology-core

Supplemental Figures



Supplemental Figure 1. Validation of hair cell detection analysis and location estimation. Whole cochlear turns **(A)** were manually annotated and evaluated with the HCAT detection analysis pipeline. Each analysis generated cochleograms **(B)**, reporting the 'ground truth' result obtained from manual segmentation (*dark lines*) superimposed onto the cochleogram generated from hair cells detected by the HCAT analysis (*light line*s). The best frequency estimation error was calculated as an octave difference of predicted best frequency for every hair cell vs their manually assigned frequency using the imageJ plugin **(C)**. Optimal cell detection and non-maximum suppression thresholds were discerned via a grid search by maximizing the true positive rate penalized by the false positive and false negative rates **(D)**. Black lines on the curves **(E)** denote the optimal hyperparameter value.



Supplemental Figure 2. Training data augmentation pipeline. Training images underwent data augmentation steps, increasing the variability of our dataset and improving resulting model performance. Examples of each transformation are shown on exemplar grids (*bottom*). Each of these augmentation steps were probabilistically applied sequentially (left to right, as shown by arrows) during every epoch.

References

1. Ashmore J. Tonotopy of cochlear hair cell biophysics (excl. mechanotransduction). *Current opinion in physiology.* 2020;18:1-6.
2. Lim DJ. Functional structure of the organ of Corti: a review. *Hearing Res.* 1986;22(1-3):117-146.
3. Urata S, Iida T, Yamamoto M, et al. Cellular cartography of the organ of Corti based on optical tissue clearing and machine learning. *eLife.* 2019;8.
4. Cortada M, Sauteur L, Lanz M, Levano S, Bodmer D. A deep learning approach to quantify auditory hair cells. *Hearing Res.* 2021;409:108317.
5. Potter PK, Bowl MR, Jeyarajan P, et al. Novel gene function revealed by mouse mutagenesis screens for models of age-related disease. *Nature communications.* 2016;7(1):1-13.
6. Kenyon EJ, Kirkwood NK, Kitcher SR, et al. Identification of a series of hair-cell MET channel blockers that protect against aminoglycoside-induced ototoxicity. *JCI Insight.* 2021;6(7).
7. Zou Z, Shi Z, Guo Y, Ye J. Object detection in 20 years: A survey. *arXiv preprint arXiv:190505055.* 2019.
8. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence.* 2016;39(6):1137-1149.
9. Ezhilarasi R, Varalakshmi P. Tumor detection in the brain using faster R-CNN. Paper presented at: 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on2018.
10. Yang S, Fang B, Tang W, Wu X, Qian J, Yang W. Faster R-CNN based microscopic cell detection. Paper presented at: 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)2017.
11. Zhang J, Hu H, Chen S, Huang Y, Guan Q. Cancer cells detection in phase-contrast microscopy images based on Faster R-CNN. Paper presented at: 2016 9th international symposium on computational intelligence and design (ISCID)2016.
12. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition2022.
13. Rasmussen CE. Gaussian processes in machine learning. Paper presented at: Summer school on machine learning2003.
14. Greenwood DD. A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America.* 1990;87(6):2592-2605.
15. Beurg M, Barlow A, Furness DN, Fettiplace R. A Tmc1 mutation reduces calcium permeability and expression of mechanoelectrical transduction channels in cochlear hair cells. *Proceedings of the National Academy of Sciences.* 2019;116(41):20743-20749.
16. Fang Q-J, Wu F, Chai R, Sha S-H. Cochlear surface preparation in the adult mouse. *JoVE (Journal of Visualized Experiments).* 2019(153):e60299.
17. Fu X, An Y, Wang H, et al. Deficiency of Klc2 induces low-frequency sensorineural hearing loss in C57BL/6 J mice and human. *Molecular Neurobiology.* 2021;58(9):4376-4391.
18. György B, Nist-Lund C, Pan B, et al. Allele-specific gene editing prevents deafness in a model of dominant progressive hearing loss. *Nature medicine.* 2019;25(7):1123-1130.
19. He Z-H, Pan S, Zheng H-W, Fang Q-J, Hill K, Sha S-H. Treatment with calcineurin inhibitor FK506 attenuates noise-induced hearing loss. *Frontiers in Cell and Developmental Biology.* 2021;9:648461.
20. Hill K, Yuan H, Wang X, Sha S-H. Noise-induced loss of hair cells and cochlear synaptopathy are mediated by the activation of AMPK. *Journal of Neuroscience.* 2016;36(28):7497-7510.
21. Kim J, Xia A, Grillet N, Applegate BE, Oghalai JS. Osmotic stabilization prevents cochlear synaptopathy after blast trauma. *Proceedings of the National Academy of Sciences.* 2018;115(21):E4853-E4860.
22. Lee S, Jeong H-S, Cho H-H. Atoh1 as a coordinator of sensory hair cell development and regeneration in the cochlea. *Chonnam Medical Journal.* 2017;53(1):37-46.

23. Li S, Mecca A, Kim J, et al. Myosin-VIIa is expressed in multiple isoforms and essential for tensioning the hair cell mechanotransduction complex. *Nature communications.* 2020;11(1):1-15.

24. Mao B, Wang Y, Balasubramanian T, et al. Assessment of auditory and vestibular damage in a mouse model after single and triple blast exposures. *Hearing Res.* 2021;407:108292.

25. Sang Q, Li W, Xu Y, et al. ILDR1 deficiency causes degeneration of cochlear outer hair cells and disrupts the structure of the organ of Corti: a mouse model for human DFNB42. *Biology open.* 2015;4(4):411-418.

26. Sethna S, Zein WM, Riaz S, et al. Proposed therapy, developed in a Pcdh15-deficient mouse, for progressive loss of vision in human Usher syndrome. *Elife.* 2021;10:e67361.

27. Wang L, Bresee CS, Jiang H, et al. Scleraxis is required for differentiation of the stapedius and tensor tympani tendons of the middle ear. *Journal of the Association for Research in Otolaryngology.* 2011;12(4):407-421.

28. Yousaf R, Meng Q, Hufnagel RB, et al. MAP3K1 function is essential for cytoarchitecture of the mouse organ of Corti and survival of auditory hair cells. *Disease Models & Mechanisms.* 2015;8(12):1543-1553.

29. Zhao X, Henderson HJ, Wang T, Liu B, Li Y. Deletion of Clusterin Protects Cochlear Hair Cells against Hair Cell Aging and Ototoxicity. *Neural Plasticity.* 2021;2021.

30. Gersten BK, Fitzgerald TS, Fernandez KA, Cunningham LL. Ototoxicity and Platinum Uptake Following Cyclic Administration of Platinum-Based Chemotherapeutic Agents. *Journal of the Association for Research in Otolaryngology.* 2020;21(4):303-321.

31. Liu W, Davis RL. Calretinin and calbindin distribution patterns specify subpopulations of type I and type II spiral ganglion neurons in postnatal murine cochlea. *Journal of Comparative Neurology.* 2014;522(10):2299-2318.

32. Kumagami H, Beitz E, Wild K, Zenner H-P, Ruppersberg J, Schultz J. Expression pattern of adenylyl cyclase isoforms in the inner ear of the rat by RT-PCR and immunochemical localization of calcineurin in the organ of Corti. *Hearing Res.* 1999;132(1-2):69-75.

33. Nagashima R, Yamaguchi T, Kuramoto N, Ogita K. Acoustic overstimulation activates 5'-AMP-activated protein kinase through a temporary decrease in ATP level in the cochlear spiral ligament prior to permanent hearing loss in mice. *Neurochemistry international.* 2011;59(6):812-820.

34. Wu P-z, Liberman MC. Age-related stereocilia pathology in the human cochlea. *Hearing Res.* 2022;422:108551-108551.

35. Kim J, Ricci AJ. In vivo real-time imaging reveals megalin as the aminoglycoside gentamicin transporter into cochlea whose inhibition is otoprotective. *Proceedings of the National Academy of Sciences.* 2022;119(9):e2117946119.

36. Jarysta A, Tarchini B. Multiple PDZ domain protein maintains patterning of the apical cytoskeleton in sensory hair cells. *Development.* 2021;148(14):dev199549.

37. Stanford JK, Morgan DS, Bosworth NA, et al. Cool otoprotective ear lumen (COOL) therapy for cisplatin-induced hearing loss. *Otology & Neurotology.* 2021;42(3):466-474.

38. Ghimire SR, Deans MR. Frizzled3 and Frizzled6 cooperate with Vangl2 to direct cochlear innervation by type II spiral ganglion neurons. *Journal of Neuroscience.* 2019;39(41):8013-8023.

39. Ghimire SR, Ratzan EM, Deans MR. A non-autonomous function of the core PCP protein VANGL2 directs peripheral axon turning in the developing cochlea. *Development.* 2018;145(12):dev159012.

40. Kim KX, Payne S, Yang-Hood A, et al. Vesicular glutamatergic transmission in noise-induced loss and repair of cochlear ribbon synapses. *Journal of Neuroscience.* 2019;39(23):4434-4447.

41. Lingle CJ, Martinez-Espinosa PL, Yang-Hood A, et al. LRRC52 regulates BK channel function and localization in mouse cochlear inner hair cells. *Proceedings of the National Academy of Sciences.* 2019;116(37):18397-18403.

42. Hayashi Y, Chiang H, Tian C, Indzhykulian AA, Edge AS. Norrie disease protein is essential for cochlear hair cell maturation. *Proceedings of the National Academy of Sciences.* 2021;118(39):e2106369118.

43. Lin T. LabelImg. In: Github; 2015.

44. Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv preprint arXiv:171105101.* 2017.

45. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data.* 2019;6(1):1-48.

46. Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. *Computing in science & engineering.* 2011;13(2):22-30.

47. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research.* 2011;12:2825-2830.

48. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:191201703.* 2019.