Cell Reports, Volume *26*

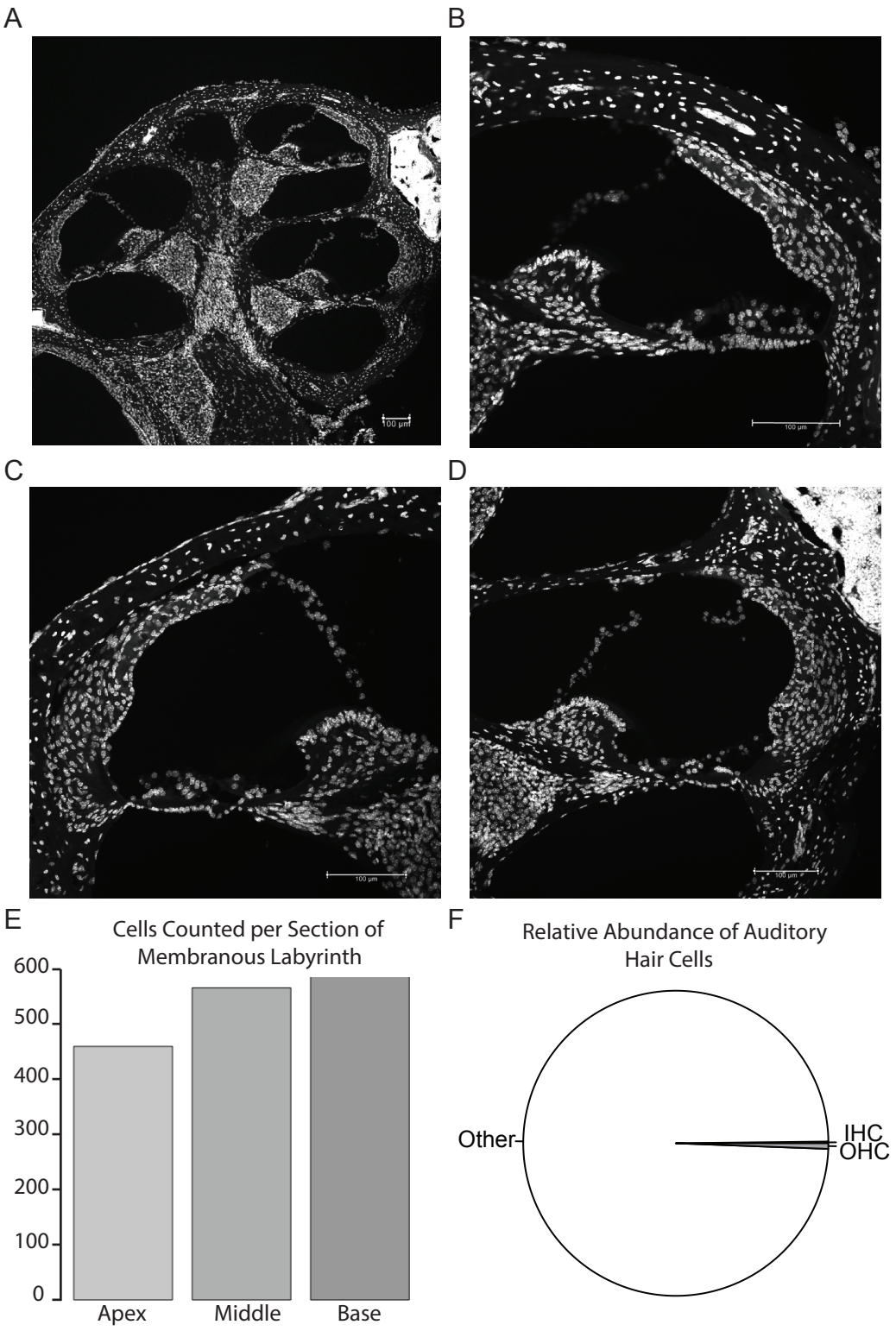
Supplemental Information

Insights into the Biology of Hearing and Deafness Revealed by Single-Cell RNA Sequencing

Paul T. Ranum, Alexander T. Goodwin, Hidekane Yoshimura, Diana L. Kolbe, William D. Walls, Jin-Young Koh, David Z.Z. He, and Richard J.H. Smith

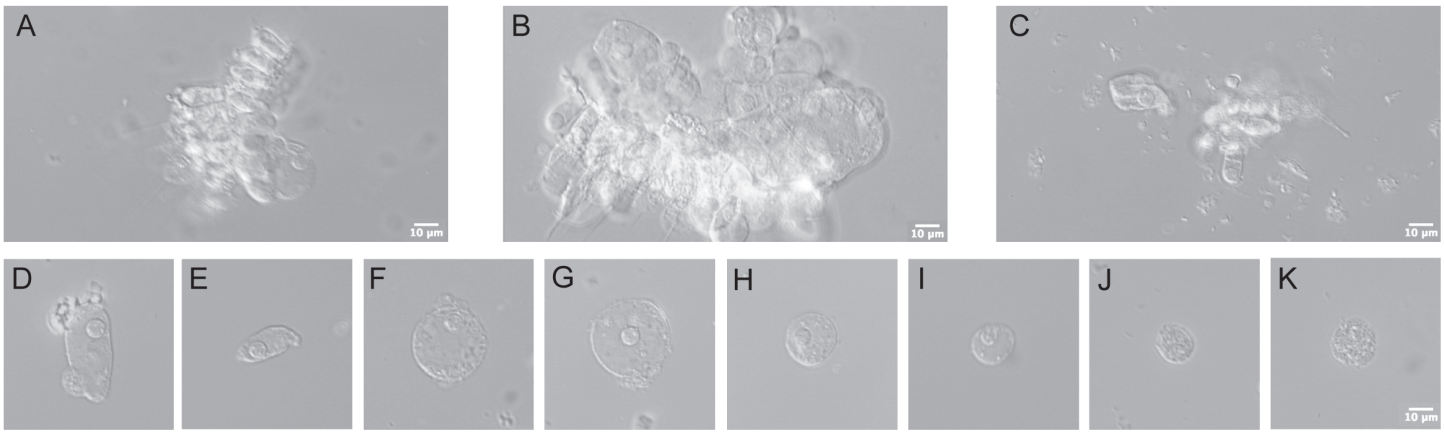
Publication	Age	Approach	Tissue	Isolation	Style	Splicing Analysis	Number of cells
Cai et al. 2015	p0	RNA-Seq	Cochlear Hair Cells	Flow	Bulk	No	NA
	p0	RNA-Seq	Cochlear Duct	Flow	Bulk	No	NA
Elkon et al. 2015	p0	RNA-Seq	Cochlear Hair Cells	Flow	Bulk	No	NA
	p0	RNA-Seq	Cochlear Epithelial (non-sensory)	Flow	Bulk	No	NA
	p0	RNA-Seq	Vestibular Hair Cells	Flow	Bulk	No	NA
	p0	RNA-Seq	Vestibular Epithelial (non-sensory)	Flow	Bulk	No	NA
Waldhaus et al. 2015	p2	PCR	OHC	Flow	Single	No	205
	p2	PCR	IHC	Flow	Single	No	40
	p2	PCR	Deiters	Flow	Single	No	178
	p2	PCR	Pillar	Flow	Single	No	186
	p2	PCR	Inner Phal	Flow	Single	No	69
	p2	PCR	Inner Border	Flow	Single	No	27
	p2	PCR	GER cells	Flow	Single	No	103
Burns et al. 2015	p1	RNA-Seq	Cochlear SC	Flow + FluidigmC1	Single	No	18
	p1	RNA-Seq	Cochlear NSC	Flow + FluidigmC1	Single	No	64
	p1	RNA-Seq	Cochlear HC	Flow + FluidigmC1	Single	No	10
	p1	RNA-Seq	Cochlear Medial SC	Flow + FluidigmC1	Single	No	NA
	p1	RNA-Seq	Cochlear Lateral SC	Flow + FluidigmC1	Single	No	NA
	p1	RNA-Seq	Utricle TEC	Flow + FluidigmC1	Single	No	NA
	p1	RNA-Seq	Utricle SC	Flow + FluidigmC1	Single	No	NA
	p1	RNA-Seq	Utricle HC	Flow + FluidigmC1	Single	No	NA
	e16.5	RNA-Seq	Cochlear Epithelium	Dissection	Bulk	No	NA
	e16.5	RNA-Seq	Vestibular Epithelium	Dissection	Bulk	No	NA
	p0	RNA-Seq	Cochlear Epithelium	Dissection	Bulk	No	NA
	p0	RNA-Seq	Vestibular Epithelium	Dissection	Bulk	No	NA
Scheffer et al. 2015	e16	RNA-Seq	GFP + Cochlear (HC)	Flow	Bulk	No	NA
	p7	RNA-Seq	GFP + Utricle (HC)	Flow	Bulk	No	NA
	e16	RNA-Seq	GFP - Cochlear	Flow	Bulk	No	NA
	p7	RNA-Seq	GFP - Utricle	Flow	Bulk	No	NA
Liu et al. 2014	Mature	Microarray	OHC	Manual	Bulk	No	NA
	Mature	Microarray	IHC	Manual	Bulk	No	NA
Li et al. 2018	1 month	RNA-Seq	OHC (3 biological replicates)	Manual	Bulk	No	~1000
	1 month	RNA-Seq	IHC (2 biological replicates)	Manual	Bulk	No	~1000
This Study	p15	RNA-Seq	OHC	Manual	Single	Yes	68
	p30	RNA-Seq	OHC	Manual	Single	Yes	5
	p70	RNA-Seq	OHC	Manual	Single	Yes	47
	p228	RNA-Seq	OHC	Manual	Single	Yes	5
	p15	RNA-Seq	IHC	Manual	Single	Yes	35
	p30	RNA-Seq	IHC	Manual	Single	Yes	1
	p70	RNA-Seq	IHC	Manual	Single	Yes	5
	p228	RNA-Seq	IHC	Manual	Single	Yes	1
	p15	RNA-Seq	DC	Manual	Single	Yes	37
	p70	RNA-Seq	DC	Manual	Single	Yes	2

Supplementary Table 1 (Related to Figure 1): Gene expression studies of specific inner ear cell types. This table broadly summarizes the experimental design, tissues, time-points, and cells profiled in previously published, cell-type-specific, gene expression studies performed using mammalian inner ear tissue.

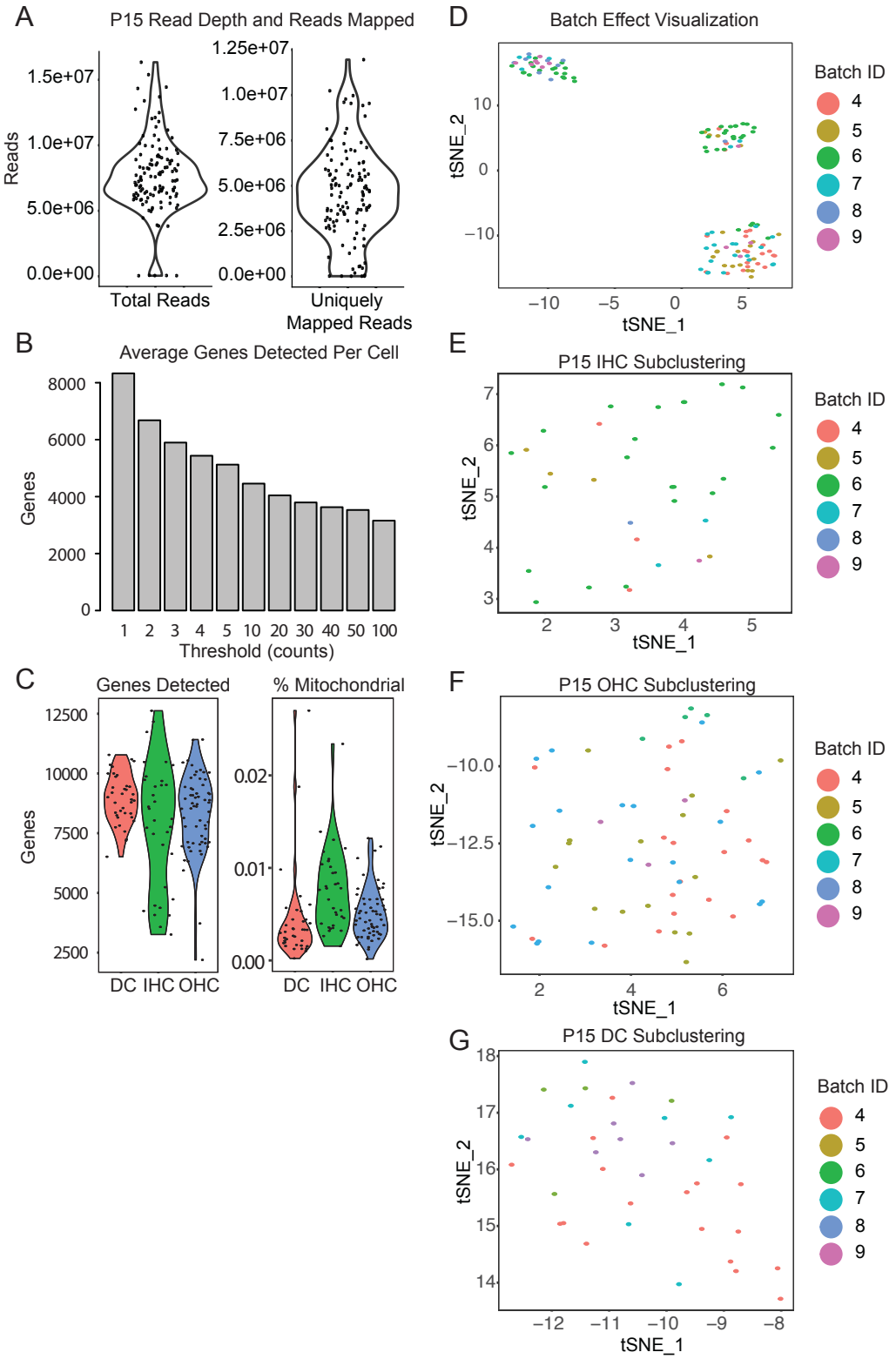


Supplementary Figure 1 (related to Figure 1): Estimation of relative hair cell abundance in the membranous labyrinth of the p15 mouse cochlea

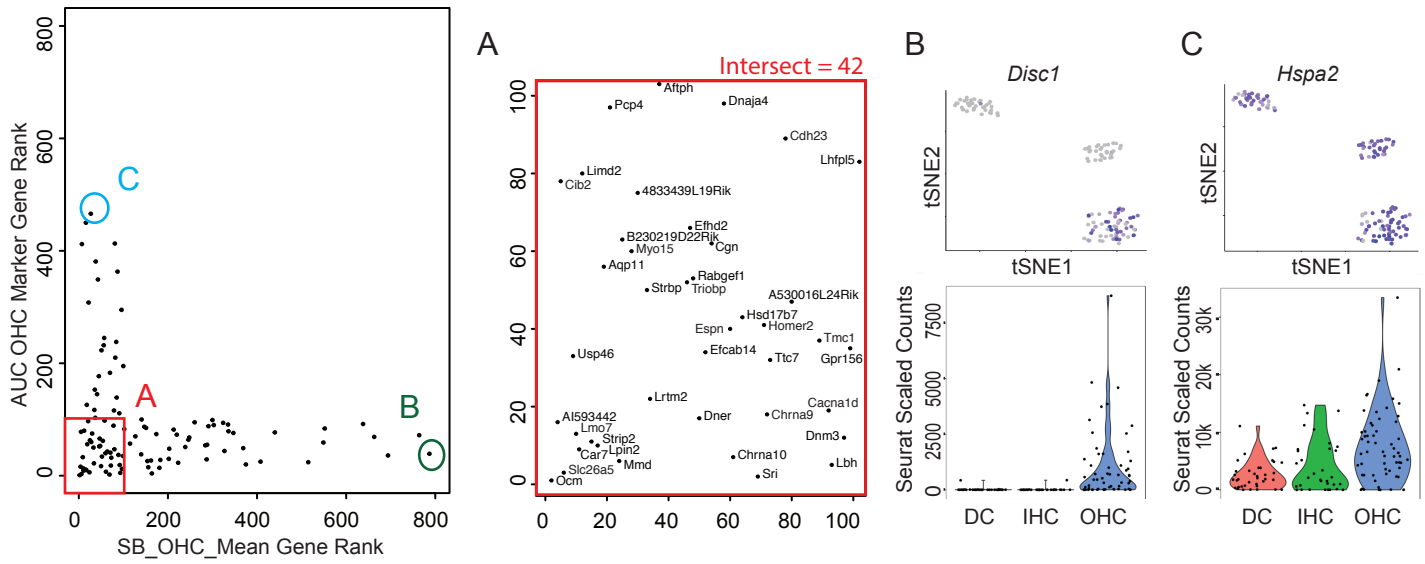
A. DAPI-stained plastic cross section of the p15 mouse cochlea. The apical (**B**), middle (**C**) and basal (**D**) turns. **E**. The number of DAPI-stained nuclei at each position was counted. **F.** Cell counting from each position was expanded to estimate the total number of sensory and non-sensory hair cells in the cochlear membranous labyrinth. The relative abundance of IHCs and OHCs to all non-sensory cells is shown.



Supplementary Figure 2 (related to Figure 1): Frequently seen cell types
Many cell types are available for single cell isolation in addition to the IHCs, OHCs and DCs targeted for these experiments. Panels **A-C** show groups of partially dissociated cells containing OHCs. In panel **A**, we can see a large round cell with a prominent nucleus located in close proximity to OHCs; there are also several smaller round cells with prominent nuclei. From the bottom, you can see DCs with their characteristic phalangeal process visible as a thin line. Panel **B** shows a wider assortment of round cells ranging from roughly 10um to 30um in diameter. The proximity of these cells to OHCs suggests that they may be either Hensen's or Claudius' cells. Panels **C-E** show cells that are more oblong. Panels **F-I** show round cell types of varying diameter similar to those seen in panels **A** and **B**. Panels **J** and **K** are round cells isolated from samples of stria vascularis. These cells have a speckled and rough morphology, and are usually less than 15µm in diameter.



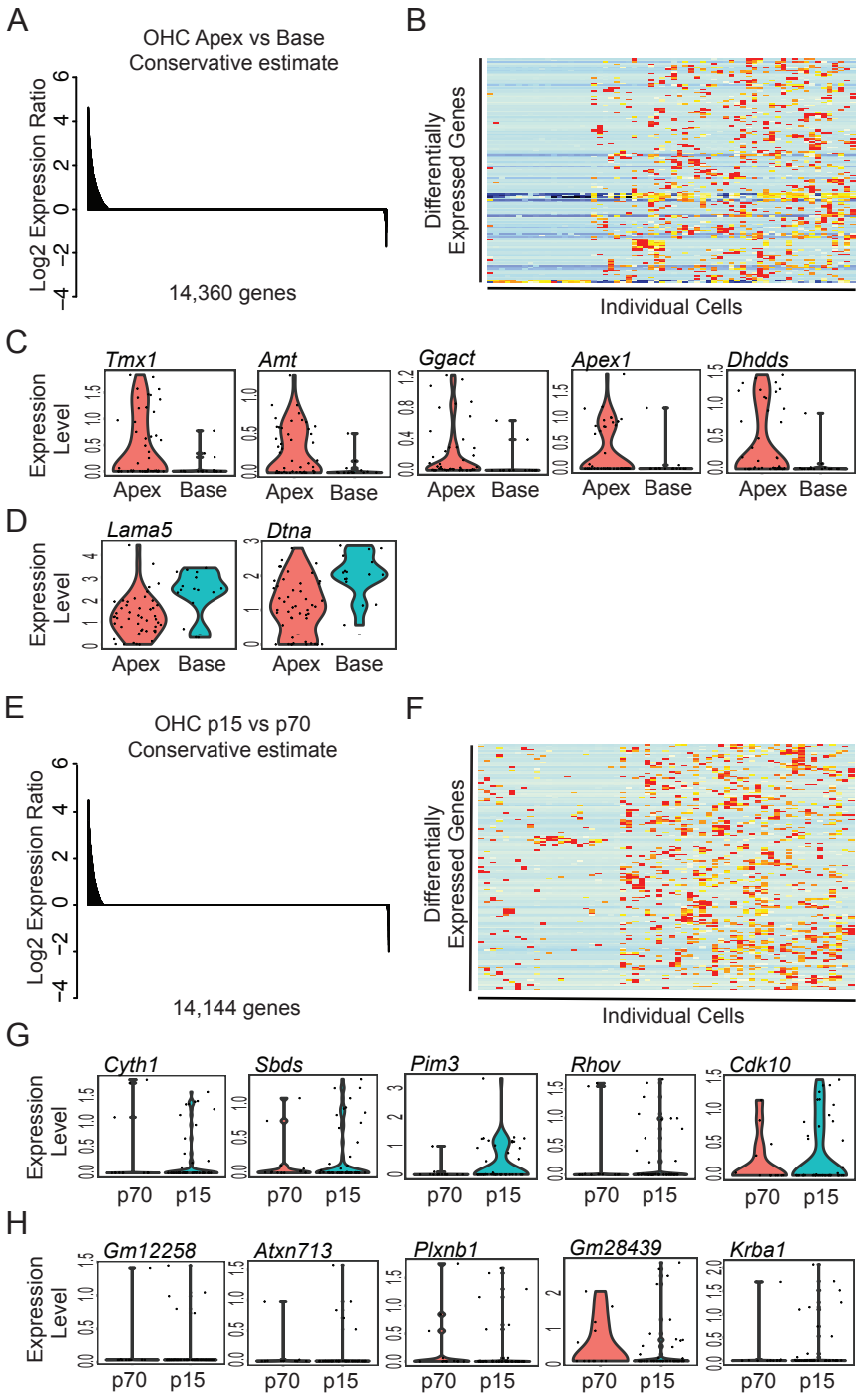
Supplementary Figure 3 (related to Figures 1 and 2): Single-cell sequencing quality metrics
Quality metrics were used to exclude poorly performing and outlier cells. **A.** The distributions of total reads per cell and uniquely mapped reads per cell are shown. 153 p15 cells from DCs, IHCs, and OHCs are included as these numbers are generated prior to filtering. **B.** The number of genes detected per cell >= the threshold value indicated between 1 and 100 Seurat scaled counts. **C.** The number of genes detected and percent mitochondrial genes detected per cell in the dataset after filtering. **D-F.** Assessment of batch effects on cluster distribution. Cells originating from six batches (sequencing runs) were included in this analysis. Each dot represents a single cell and is colored by batch. Note the minimal impact of batch assignment on clustering (**D**) and subclustering positions (**E-F**).



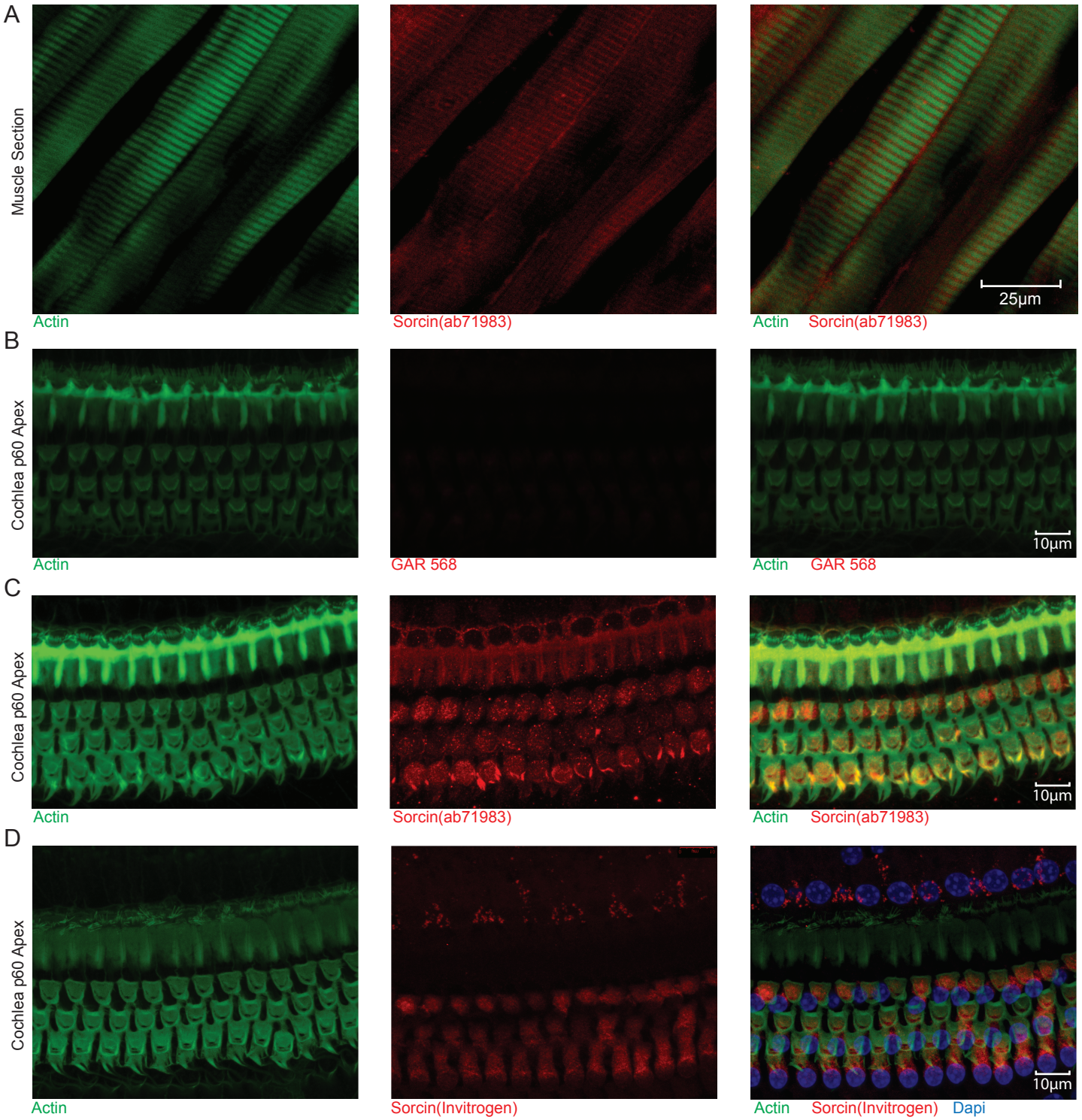
Supplementary Figure 4 (related to Figure 2): Explanation of AUC classifier use for detection of cell-type defining gene expression

Rankings made using gene expression levels (TPM or counts) differ from rankings made using Seurat's Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) classifier. In rankings made using gene expression levels, the #1 ranked gene represents the gene that is detected with the highest normalized TPM or counts value. In rankings made using the AUC classifier, the #1 ranked gene represents the gene that has the highest true-positive-to-false-positive ratio. When comparing clusters of single cells, the #1 ranked cluster-defining gene is the gene that has the highest probability of being detected within the cells in that cluster and the least probability of being detected in any other cluster. Both metrics are useful, but in different ways. Expression levels define the genes that are most highly expressed in a dataset; the AUC classifier identifies genes that have the highest probability of being detected in a given cluster across a range of expression levels. As such, the AUC classifier can be used to identify 'cluster-defining genes'.

To illustrate the utility of the AUC classifier to detect cluster-defining genes across a range of expression levels, we have compared the top 100 genes ranked by the AUC classifier to the top 100 genes ranked by gene expression level (Seurat scaled counts) using a simulated bulk dataset (SB_OHC_Mean). This comparison is visible in the main panel on the left. **A.** The intersect of these two lists (defined by the red square labeled A) contains 42 genes, which represents the number of genes ranked within the top 100 genes by both gene expression level and the AUC classifier. **B.** *Disc1*, circled in green in the main panel, is an example of a gene ranked high (39th) by the AUC classifier but low (787th) by gene expression level. When evaluated by a clustering heatmap (colored dots represent expression), the specificity (upper panel of B) of the expression within the OHC cluster is striking. From the violin plots (bottom panel of B), we see that the expression level of *Disc1* in OHCs is unremarkable, which is consistent with its 787th place ranking. **C.** *Hspa2*, circled in blue in the main panel on the left, is an example of a gene ranked low (466th) by the AUC classifier but high (27th) by gene expression level. When evaluated by a clustering heatmap (upper panel of C), we can see that expression is common to DC, IHC and OHC clusters. From the violin plots (lower panel of C), we see that the expression level of *Hspa2* are robust in many cells across all three cell types.

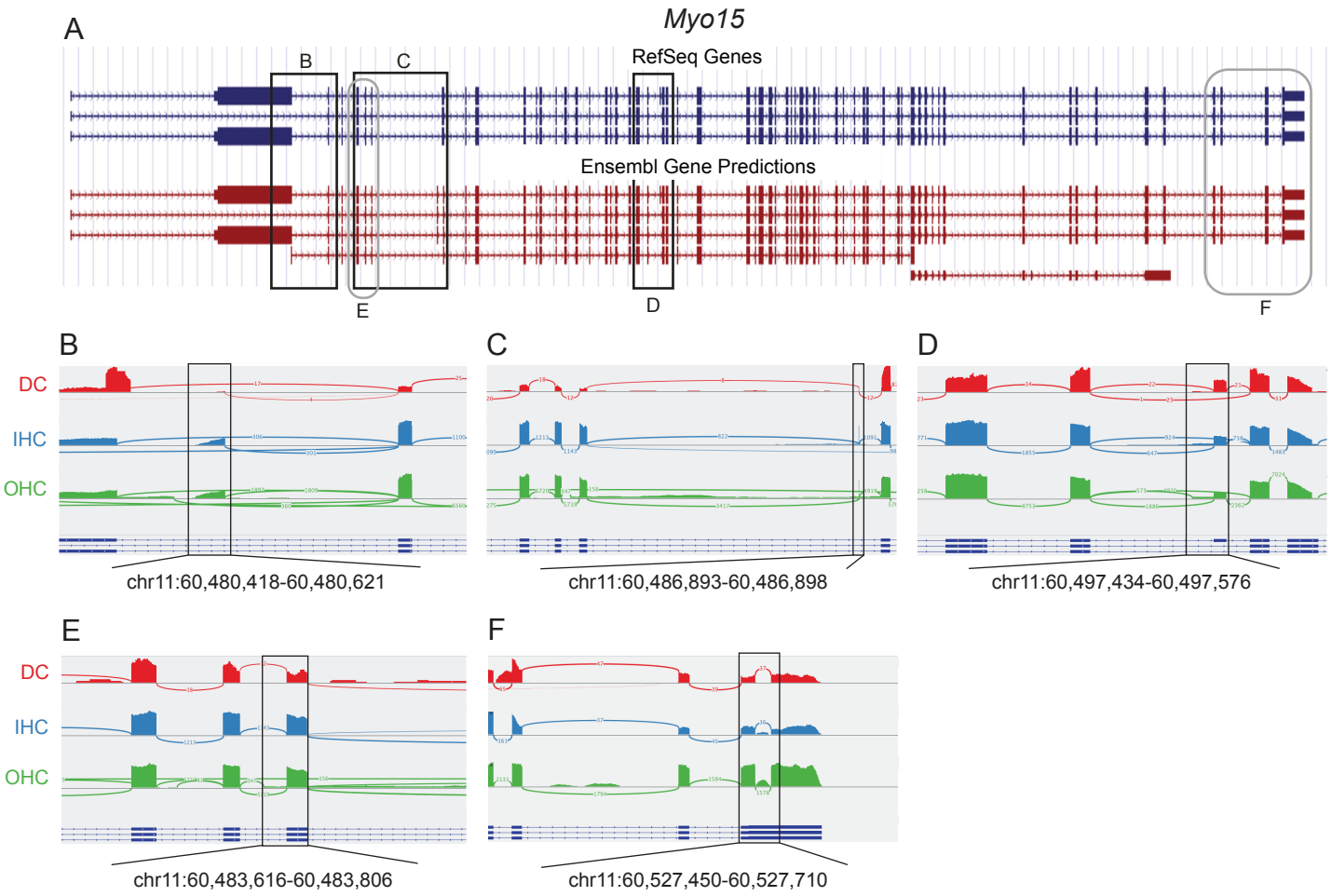


Supplementary Figure 5 (related to Figure 2): OHC differential expression between tonotopic positions and timepoints. **A.** 14,360 genes detected across 64 cells met our threshold for inclusion. 1,070 genes were conservatively estimated by SCDE to be differentially expressed among the Apex (n=48) and Base (n=16) groups. The majority (952) of these differentially expressed genes were enriched in the Apex group (Log2 Expression Ratio >0) and the remaining 118 genes were enriched in the Base group (Log2 Expression Ratio <0). **B.** Of these differentially expressed genes, 155 meet the threshold for statistical significance after multiple testing correction (p value < .05). 153 genes were enriched in the Apex while only 2 were enriched in the Base. **C.** Violin plots depicting the top 5 most differentially expressed genes from the Apex, ranked by log2 fold change. **D.** The two differentially expressed genes from the Base group meeting the criteria for statistical significance. **E.** 14,144 genes detected across 61 cells met our threshold for inclusion. 1,227 genes were conservatively estimated by SCDE to be differentially expressed between the p15 (n=48) and p70 (n=13) groups. The majority (1151) of these differentially expressed genes were enriched in the p15 group, while the remaining 76 genes were enriched in the p70 group. **F.** None of the 1151 genes met the threshold for statistical significance (p value <.05) after multiple testing correction. Genes with SCDE conservative estimate scores of >3 and <-1 were included in this heatmap, which represent the most differentially expressed genes in this dataset. **G.** The top 5 genes with the highest conservative estimate scores are shown representing the genes most enriched at the p15 timepoint. **H.** The bottom 5 genes with the lowest conservative estimate scores are shown representing the genes most enriched at the p70 timepoint.



Supplementary Figure 6 (related to Figure 3): Sorcin localizes to OHCs

A. Control: muscle sections stained with phalloidin (green) and anti-sorcin antibody (ab71983; Abcam, Cambridge, MA) (red). The alternating phalloidin- and sorcin-labeled bands suggest decoration of sarcomere I and A bands, respectively. **B.** Secondary-only control to identify background fluorescence of cochlear tissues stained with goat anti-rabbit (GAR) 568 secondary antibody. **C.** Apical cochlear whole mount labeled with rabbit anti-sorcin ab71983 antibody. Distinct labeling of cell bodies of OHCs and the apical surface of pillar cells is observed. Circular staining around the cell bodies of IHCs is also visible, as is bright labeling of DC-OHC junctions in the third row of OHCs. Based on this staining pattern, we determined Sorcin (ab71983) had substantial background. **D.** Apical cochlear whole mount labeled with rabbit anti-sorcin antibody from Invitrogen (Carlsbad, CA) shows more specific and robust staining of the cell bodies of OHCs. Punctate staining is seen along the row of IHCs and may be attributed to either IHCs or supporting cells. Both antibodies robustly label the OHC cell body and lateral membrane.



Supplementary Figure 7 (related to Figure 5): Splicing analysis of mouse Myo15
A. Overview of mouse *Myo15* showing annotated isoforms in the RefSeq and Ensembl databases visualized using the UCSC genome browser. Black boxes indicate published structural feature landmarks at the indicated positions (B-D). Grey rounded boxes indicate the regions shown in panels E-F, which contain unannotated *Myo15* features. **B.** An alternative transcription start site located after the large canonical exon 2. **C.** A previously reported 6bp exon annotated in Ensembl but not in RefSeq. **D.** A previously reported alternative splice acceptor site, which would result in the inclusion of a premature stop. **E.** A unannotated alternative splice acceptor site the inclusion of which would result in a premature stop. **F.** A unannotated splice site in the 3' UTR.