Supporting online material for:

Mapping the regulatory landscape of auditory hair cells from single-cell multi-omics data

Shuze Wang[1,2], Mary P. Lee[1], Scott Jones[1], Jie Liu[2], Joerg Waldhaus[1]

Corresponding author: joergwal@med.umich.edu

This document includes Supplemental Tables, Supplemental Methods, Supplemental Figures, and Supplemental References

SUPPLEMENTAL METHODS

Supplemental Table S1

GENE SYMBOL	DEVELOPMENT RELEVANCE	LITERATURE REFERENCE
Stat5b	IGF signaling	(Hertzano et al. 2007)
Rbpj	Notch signaling mediator	(Basch et al. 2011)
Sox4	SOXC TF	(Gnedeva and Hudspeth 2015)
Bach1	TGFβ signaling in SCs	(Cheng et al. 2019)
Nfatc4	Apoptosis; damage independent expression	(Zhang et al. 2019)
Irf3	Constitutively expressed immune gene	(Cai et al. 2014)
Srebf2	TGFβ signaling in SCs	(Cheng et al. 2019)
Hmga2	Stemness	(Golden et al. 2015)
Klf4	Yamanaka factor	(Lopez-Juarez et al. 2019)
Atf3	**Expression in response to noise**	(Maeda et al. 2020)
Esr1	Estrogen signaling	(Simonoska et al. 2009)
Etv5	FGF signaling	(Ebeid and Huh 2020)
Isl1	Prosensory development	(Radde-Gallwitz et al. 2004)
Hey2	FGF signaling	(Doetzlhofer et al. 2009)
Nr4a3	SC development	(Maass et al. 2016)
Gata3	Prosensory development	(Luo et al. 2013)
Rora	SC development	(Maass et al. 2016)

Jun	SC development	(Sanz et al. 1999)
Hif1a	**Expression in response to noise**	(Chung et al. 2004)
Tcf7l2	WNT signaling	(Huang et al. 2011)
Rest	Development	(Wilkerson et al. 2019)
Rel	SC development	(Scheffer et al. 2015)
Zbtb7a	**Expression in response to noise**	(Yang et al. 2015)
Arntl	Embryonic development	(Perl et al. 2018)
Mtf1	GATA3 target	(Alvarado et al. 2009)

Supplemental Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MYOSINVIIA	Proteus Biosciences, 25-6790	Cat # 25-6790; RRID:AB_2314840
SOX2	Santa Cruz	Cat# sc-17320; RRID:AB_2286684
TCF4	Proteintech	Cat# 13838-1-AP; RRID:AB_2199812
GLIS3 Polyclonal Antibody	Thermo Scientific	Cat# PA5-41677; RRID:AB_2606554

FOXO4	Proteintech	Cat # 21535-1-AP; RRID:AB_10732960
ZNF238 Polyclonal Antibody	Thermo Fisher Scientific	Cat# PA5100372; RRID:AB_2849885
Chemicals, Peptides, and Recombinant Proteins		
Tamoxifen	Sigma	Cat# T5648
DAPI	Sigma	Cat# 10236276001
O.C.T compound	Fisher Scientific	Cat# 4585
SYTOX Red Dead Cell Stain	Thermo Fisher Scientific	Cat# S34859
Accutase	Innovative Cell Technologies	Cat# AT104
Thermolysin from Geobacillus stearothermophilus	Sigma	Cat# P1512
16% PFA	Electron Microscopy Sciences	Cat# 15710
Triton-X	Sigma	Cat# 1086431000
Corn Oil	Sigma	Cat# C8267
1XPhosphate Buffered Saline	Thermo Fisher Scientific	Cat# 200120027
Bovine Serum Albumin	Thermo Fisher Scientific	Cat# BP9700100

Critical Commercial Assays		
Chromium Next GEM Chip G Single Cell Kit	10x Genomics	Cat# PN-1000120
Chromium Next GEM Single Cell 3' Kit v3.1	10x Genomics	Cat# PN-1000268
Library Construction Kit	10x Genomics	Cat# PN-1000190
Dual Index Kit TT Set A	10x Genomics	Cat# PN-1000215
10x Chromium and Next GEM accessory kit	10x Genomics	Cat# PN-100202
Chromium Next GEM Chip H Single Cell Kit	10x Genomics	Cat# PN-1000162
Chromium Next GEM Single Cell ATAC Library and gel bead kit v 1.1	10x Genomics	Cat# PN-000175
Single Index kit N Set4 A	10x Genomics	Cat# PN-1000212
RNAScope Multiplex Fluorescent Reagent Kit	Advanced Cell Diagnostic	Cat# 323110
mm-*Pkhd1l1*	Advanced Cell Diagnostic	Cat# 44001-C3
Deposited Data		

Single-cell RNA-seq from mouse cochlea at P2	This paper	GEO under accession number GSE157398
Single-cell ATAC-seq from mouse cochlea at P2	This paper	GEO under accession number GSE157398
Experimental Models: Organisms/Strains		
ATOH1-GFP	Jackson Laboratory	IMSR Cat# JAX:013593, RRID:IMSR_JAX:013593
FGFR3-iCre	Jackson Laboratory	IMSR Cat# JAX:025809, RRID:IMSR_JAX:025809
Ai14-tdTomato	Jackson Laboratory	IMSR Cat# JAX:007914, RRID:IMSR_JAX:007914
Software and Algorithms		
CellRanger (v. 3.0.0)	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
CellRangerATAC (v. 1.1.0)	10x Genomics	https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/installation
SAMtools (v. 1.5)	(Li et al. 2009)	http://www.htslib.org/;

		RRID:SCR_002105
BEDTools (v. 2.29.1)	(Quinlan and Hall 2010)	https://bedtools.readthedocs.io/en/latest/; RRID:SCR_006646
MACS2 (v. 2.1.2)	(Zhang et al. 2008)	https://github.com/macs3-project/MACS; RRID:SCR_013291
ataqv (v. 1.0.0)	(Orchard et al. 2020)	https://github.com/ParkerLab/ataqv
FIMO (v. 5.1.0)	(Grant et al. 2011)	http://meme-suite.org/tools/fimo
HINT-ATAC (v. 0.12.3)	(Li et al. 2019)	https://www.regulatory-genomics.org/
LiftOver	(Kent et al. 2002)	https://genome.ucsc.edu/cgi-bin/hgLiftOver; RRID:SCR_018160
Integrative Genomics Viewer (v. 2.4.2)	(Robinson et al. 2011)	http://software.broadinstitute.org/software/igv/; RRID:SCR_011793
SnapTools (v. 1.4.8)	(Fang et al. 2019)	https://github.com/r3fang/SnapTools; RRID:SCR_018097
bedGraphToBigWig (v. 4)	(Kent et al. 2010)	http://hgdownload.soe.ucsc.edu/a

		dmin/exe/linux.x86_64.v369/
diffTF (v. 1.6)	(Berest et al. 2019)	https://git.embl.de/grp-zaugg/diffTF
DAVID (v. 6.8)	(Huang et al. 2009)	https://david.ncifcrf.gov/; SCR_001881
R (v. 3.5.1)	R core	https://www.r-project.org/
Bioconductor	(Huber et al. 2015)	http://www.bioconductor.org/; RRID: SCR_006442
Seurat (v. 3.1.2)	(Stuart et al. 2019)	https://satijalab.org/seurat/install.html; RRID:SCR_007322
LIGER (v. 0.5.0)	(Welch et al. 2019)	https://github.com/welch-lab/liger; RRID:SCR_018100
SnapATAC (v. 1.0.0)	(Fang et al. 2019)	https://github.com/r3fang/SnapTools; RRID:SCR_018097
chromVAR (v. 1.4.1)	(Schep et al. 2017)	http://www.bioconductor.org/packages/release/bioc/html/chromVAR.html
GENIE3 (v. 1.6.0)	(Huynh-Thu et al. 2010; Aibar et al. 2017)	https://bioconductor.org/packages/release/bioc/html/GENIE3.html; RRID:SCR_000217

AUCell (v. 1.4.1)	(Aibar et al. 2017)	http://www.bioconductor.org/packages/release/bioc/html/AUCell.html; RRID:SCR_017247
Cicero (v. 1.0.14)	(Pliner et al. 2018)	https://cole-trapnell-lab.github.io/cicero-release/
CellTrails (v. 1.0.0)	(Ellwanger et al. 2018)	https://bioconductor.riken.jp/packages/3.8/bioc/html/CellTrails.html
Slingshot (v. 1.8.0)	(Street et al. 2018)	https://www.bioconductor.org/packages/release/bioc/html/slingshot.html; RRID:SCR_017012
Monocle (v. 2.10.1)	(Trapnell et al. 2014; Qiu et al. 2017a; Qiu et al. 2017b)	http://cole-trapnell-lab.github.io/monocle-release/docs/; RRID:SCR_016339
ChIPseeker (v. 1.18.0)	(Yu et al. 2015)	https://bioconductor.riken.jp/packages/release/bioc/html/ChIPseeker.html
GenomicRanges (v. 1.34.0)	(Lawrence et al. 2013)	https://bioconductor.riken.jp/packages/release/bioc/html/GenomicRanges.html; RRID:SCR_000025
TxDb.Mmusculus.UCSC.mm10.knownGene (v. 3.4.4)	(Bioconductor Core Team 2018)	https://bioconductor.riken.jp/packages/release/data/annotation/html/TxDb.Mmusculus.UCSC.mm10.k

		nownGene.html
org.Mm.eg.db (v. 3.7.0)	(Carlson 2018)	https://bioconductor.riken.jp/packages/release/data/annotation/html/org.Mm.eg.db.html
HOCOMOCO v10 database	(Kulakovskiy et al. 2013)	https://hocomoco10.autosome.ru/; RRID:SCR_005409
JASPAR 2020 database	(Fornes et al. 2020)	http://jaspar.genereg.net/; RRID:SCR_003030
Other		
MoFlo Astrios Cell Sorter	Beckman Coulter	RRID:SCR_018893
Leica SP8 LIGHTNING confocal microscope	Leica	RRID:SCR_018169
10x Chromium and Next GEM accessory kit	10x Genomics	Cat# 1000202

Supplemental Methods

Experimental Model

All animal maintenance and experimental procedures were performed in accordance with NIH guidelines and were approved by the Institute Animal Care and Use Committee at the University of Michigan (protocol: PRO00008096, JW).

All mice were maintained on a mixed genetic background. FGFR3-iCre (Young et al. 2010) were crossed with Ai14-tdTomato (Madisen et al. 2010) and ATOH1-GFP (Rose et al. 2009). In order to induce FGFR3-iCre activity, P0 pups were given intraperitoneal injections of tamoxifen (Sigma, T5648-1G) at 200 mg/kg, diluted in corn oil (Sigma, C8267). Cochlear samples were isolated at P2 from FGFR3-iCre;Ai14-tdTomato;ATOH1-GFP-mice.

Single Cell Isolation and Flow Sorting

At P2 the cochlear ducts of FGFR3-iCre;Ai14-tdTomato;ATOH1-GFP pups were microdissected and divided into apical and basal samples. Tissues were then dissociated with enzymatic (accutase and thermolysin) and mechanical dissociation and passed through a strainer in order to create a single-cell suspension, as previously described (Durruthy-Durruthy et al. 2014). In order to enrich for HCs and PC/DCs before sequencing, cells were purified with FACS using a MoFlo Astrios (Beckman Coulter, University of Michigan Flow Cytometry Core). The FACS gating strategy was designed to isolate and pool together cells expressing, tdTomato$^+$, tdTomato$^+$/GFP$^+$, and GFP$^+$ from either the apical or basal samples before being subjected to standard 10x Genomics preparations for scRNA-seq or scATAC-seq experiments.

RNAScope and Immunofluorescence Combined Staining

Cochlear samples from P2 neonatal mice were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15710) diluted in 1XPBS (Thermo, 20012017) for 3 hours at room temperature, washed in 1XPBS (Thermo), subjected to a sucrose gradient, then embedded in O.C.T. Compound (Fisher, 4585) for cryosectioning. 18 μm sections were utilized for RNAScope and immunofluorescent experiments.

For RNAScope, sections were incubated in RNase/DNase Free Water (Invitrogen, AM9932) for 5 minutes, dehydrated using an ethanol gradient (50%, 70%, 90%, 100%), and subjected to RNAScope staining using the Advanced Cell Diagnostics (ACD) RNAScope Multiplex Fluorescent Reagent Kit v2

(ACD, 323110) with the following modifications. Antigen retrieval was performed at 65°C for 5 minutes and no blocking steps were performed following development of fluorophores.

RNAScope samples were then prepared for immunofluorescence staining by washing in 0.1% Triton-X (Sigma, 1086431000) in 1XPBS (Thermo), primary antibodies were applied in 0.05% Triton-X in 1XPBS overnight at 4°C. Then, secondary antibodies were applied in 0.05% Triton-X in 1XPBS for 60 minutes followed by DAPI (Sigma, 10235676001) nuclear staining for 10 minutes. All washes between primary and secondary antibody application were performed with 1XPBS (Thermo). The following primary antibodies were used: MYOSINVIIA (Proteus Biosciences, 25-6790) and SOX2 (Santa Cruz, sc-17320). The RNAScope probe was designed by and ordered from ACD, mm-*Pkhd1l1* (44001-C3). All fluorescent images were acquired using a laser scanning confocal microscope (Leica, SP8).

Immunofluorescence Staining

Cochlear samples from P2 neonatal mice were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) diluted in 1XPBS (Thermo) for 2 hours on ice, washed in 1XPBS (Thermo), subjected to a sucrose gradient, then embedded in O.C.T. Compound (Fisher) for cryosectioning at 18 μm thickness. Sections were blocked in 1% bovine serum albumin (BSA, Thermo Scientific, BP9700100) in 0.2% Triton-X/1XPBS for 60 minutes at room temperature and primary antibodies were applied overnight at 4°C in 0.5% BSA/0.2% Triton-X/1XPBS. The following primary antibodies were used: MYOSINVIIA (Proteus Biosciences), SOX2 (Santa Cruz, sc-17320), TCF4 (Proteintech Group, 50560760), GLIS3 (Thermo Scientific, PIPA541677), FOXO4 (Proteintech Group,501733469), and ZBTB18 (Invitrogen, PA5100372). Secondary antibodies were applied in 0.5% BSA/0.2% Triton-X/1XPBS for 90 minutes at room temperature followed by DAPI (Sigma) nuclear staining. All washes following primary and secondary antibody application were performed with 0.2% Triton-X/1XPBS. All fluorescent images were acquired using a laser scanning confocal microscope (Leica, SP8).

10x Genomics Pipeline

Single-cell processing and next-generation sequencing was carried out in the Advanced Genomics Core at the University of Michigan. Sequencing was performed on the 10x Chromium and Next GEM accessory kit (10x Genomics, 1000202) using Chromium Next GEM Chip G Single Cell Kit (10x Genomics, 1000120) for scRNA-seq and Chromium Next GEM Chip H Single Cell Kit (10x Genomics, 1000162) for scATAC-seq. The following kits were utilized for library preparation: Chromium Next GEM Single Cell 3' Kit v3.1 (10x Genomics, 1000268), Library Construction Kit (10x Genomics, 1000190), and Dual Index Kit TT Set A (10x Genomics, 1000215) for scRNA-seq, and Chromium Next GEM Single Cell ATAC Library and gel bead kit v 1.1 (10x Genomics, 1000175) and Single Index Kit N Set 4 A (10x Genomics, 1000212) for scRNA-seq.

scATAC-seq Analysis

We started with preprocessing the data using SnapTools (Fang et al. 2019) (*snaptools snap-pre --min-flen=50 --min-cov=100 --max-num=20000 --keep-single=FALSE --overwrite=True*). Then we binned the whole genome into 5kb fixed windows and estimated the read coverage for each bin to generate a cell-by-bin matrix (*snaptools snap-add-bmat --bin-size-list 5000*). Bins that overlapped with ENCODE-defined blacklist regions were removed and bins within unwanted chromosomes, like ChrUn, Chr_random, and chrM, were filtered out as well. To select high-quality cells, we kept cells with a $\log_{10}$ UMI count between 3 and 5 and with a promoter ratio ranging from 0.2 to 0.8. After stringent quality control, 1,210 single cells were subject to further analysis. First, we binarized the cell-by-bin matrix. We then followed the pipeline to run the dimensionality reduction method, Diffusion maps, and selected the first 15 significant components (*eigs.dim=15*). The 15 significant components were used to construct the *k*-nearest neighbor graph and the Leiden algorithm was leveraged for clustering with resolution 0.2 (*resolution=0.2*). Finally, 6 clusters from the scATAC-seq dataset were identified and visualized by projecting meta-data, like read depth and library ID, using UMAP.

Upon clustering, cells from each cluster were aggregated to generate an ensemble track for peak calling. Peak calling was performed for each cluster separately using *runMACS* function in SnapATAC with following parameters: *--nomodel --shift 100 --ext 200 --qval 5e-2 -B --SPMR*. Two output files for each cluster were generated: a narrawPeak file including identified peaks; a bedGraph (.bdg) file was converted into BigWig format (.bw) using bedGraphToBigWig downloaded from UCSC Genome Browser for visualization. Next, all peaks were merged across clusters and a cell-by-peak matrix was created for differential analysis. Differentially accessible regions (DARs) for 6 clusters were determined using *findDAR* function in the SnapATAC with following parameters: *cluster.neg = NULL, cluster.neg.method="knn", test.method="exactTest", bcv=0.1*). In total, we identified 67,415 DARs under the adjusted *P*-value with Bonferroni correction of 0.05. Additionally, we annotated the DARs to the nearest gene by employing the *annotatePeak* function in ChIPseeker R package.

Quality Control of scATAC-seq Dataset

We used ataqv (Orchard et al. 2020), an ATAC-seq QC and visualization tool, to measure and estimate quality of the scATAC-seq data. Once 6 clusters were identified from scATAC-seq dataset, we aggregated the cells from the same cluster and treated each cluster as a small bulk ATAC-seq dataset. We included mm10 blacklist and mm10 TSS reference files and ran ataqv for each cluster separately with the following parameters: *--ignore-read-groups, --tss-extension 2000bp*. To visualize the TSS enrichment based on the transposition activity around TSS, we generated a 100bp window to aggregate the signals and took an average. Next, we applied a natural spline method to interpolate data points. Similarly, we created a 20bp window and applied the natural spline method to visualize the fragment length distribution.

scRNA-seq Analysis

The scRNA-seq dataset was analyzed using Seurat v3 pipeline (Stuart et al. 2019). We selected the cells with the number of features ranging from 600 to 8,000, and the maximum allowed fraction of mitochondrial genes per cell was 10%. Overall, 695 cells passed the quality control for further analysis. After the preprocessing step, log normalization was performed, and the top 2000 highly variable genes were identified using method *vst* with default settings. We scaled the datasets to avoid the domination of highly expressed genes and used it as input for PCA to reduce dimensions. The first 10 principal components were chosen to construct the shared nearest neighbor graph with 20 nearest neighbors (*k.param=20*). Leiden algorithm was performed to identify clusters with resolution 0.5 (*resolution=0.5*) and 11 clusters were resolved. We leveraged UMAP to visualize the scRNA-seq clustering results and meta-data information, like read depth and library ID.

To determine cell identities for each cluster, we first identified differentially expressed genes (DEGs) for each cluster using *FindAllMarkers* function in the Seurat package with following parameters: *only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, test.use = "wilcox"*. In total, we determined 5,772 DEGs for 11 clusters under the adjusted *P*-value with Bonferroni correction of 0.05. Next, we annotated cell identities by comparing cluster specific DEGs with published canonical marker genes.

Cell Type Identification in scATAC-seq Clusters using Jaccard Index Similarity Matrix

We generated a Jaccard index similarity matrix by calculating normalized overlaps between DEGs and annotated DARs from scRNA-seq and scATAC-seq data. Specifically, we performed the calculation for each comparison between pairs of scRNA-seq and scATAC-seq clusters:

$$J(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|} \quad i = 1,2,\dots,11 \;\; j = 1,2,\dots,6$$

where A_i is the DEGs of cluster i $(i = 1,2,\dots,11)$ from scRNA-seq, and B_j is the annotated DARs of cluster j $(j = 1,2,\dots,6)$ from scATAC-seq. $J(A_i, B_j)$ is the Jaccard index by calculating the number of intersected genes between A_i and B_j over the total number of unique genes of A_i and B_j. The values were centered and scaled in scATAC-seq clusters. Based on the Jaccard index similarity matrix, we annotated the scATAC-seq clusters with the most similar cell identities from scRNA-seq data.

LIGER Multi-omics Integration

We applied LIGER (Welch et al. 2019) to jointly define cell identities by leveraging scRNA-seq and scATAC-seq datasets. Briefly, LIGER delineates shared and dataset-specific features to integrate single-cell multi-omics data using integrative non-negative matrix factorization. In order to run LIGER, we first annotated the fragments to the gene level in SnapATAC and generated gene-by-cell matrix for scATAC-seq. We took gene-by-cell raw counts from scRNA-seq as another input. Next, we extracted differentially expressed genes across cell types, identified from Seurat package from scRNA-seq, as variable genes in LIGER. Then, joint matrix factorization algorithm was performed on the normalized and scaled scRNA-seq and scATAC-seq data with defined 20 latent variables ($k = 20$). We conducted quantile normalization, Louvain clustering, and dimensionality reduction and visualization using UMAP (*distance="cosine", n_neighbors=30, min_dist=0.3*). To validate the similarity-based approach, we projected the cell identities as determined from Jaccard index similarity matrix onto the co-embedding UMAP.

TF Motif Activity Estimation

We applied chromVAR (Schep et al. 2017) to infer TF-associated accessibility and to characterize potential TFs regulating cell differentiation and function from scATAC-seq dataset. Briefly, chromVAR aggregates peaks that share a common feature (e.g. TF motif) and corrects bias in terms of technical confounders (e.g. GC content, average accessibility) based on background peak sets. The accessibility of a TF motif was determined by calculating standardized z-scores. To prepare for chromVAR, the mouse HOCOMOCO v10 database was applied and peaks with less than 3 fragments overlapped (*min_fragments_per_peak = 3*) were filtered. We ran the pipeline using the default settings. A TF-by-cell matrix of z-scores was generated for further analysis.

To validate the clustering consistency of the scATAC-seq dataset between SnapATAC and chromVAR methods, we applied the dimensionality reduction and visualization method, UMAP, to the TF-by-cell matrix generated from chromVAR. SnapATAC cluster IDs were projected onto the z-score-based UMAP.

Transcription Factor Classification into Activators and Repressors

We developed an integrative approach to classify the mode of action of developmental TFs into activators and repressors by comparing two populations that share a common progenitor using multi-omics single-cell datasets. In the first step, we conducted Student *t*-tests with Bonferroni correction (*P*-adjust < 0.05) and identified differentially expressed TF genes and differential TF motifs from scRNA-seq and scATAC-seq datasets, respectively. Additionally, the Wilcoxon sum rank test and median value comparison were also provided for consideration. Classification was based on the relation between the TF gene expression level and TF motif accessibility for each individual TF.

The same pipeline was employed to identify TFs regulating HC vs PC/DC differentiation and IHC vs OHC differentiation. For the IHC/OHC comparison, Wilcoxon sum rank tests with Bonferroni correction (*P*-adjusted < 0.05) were conducted because of the biased sample size and violation of normality assumption.

To validate our classification approach, we adopted previously published algorithm, diffTF (Berest et al. 2019), and compared the overlaps between the two methods. diffTF, a bulk-based approach, compares two groups with multiple biological replicates using Pearson correlation by leveraging matching RNA-seq and ATAC-seq datasets. To apply diffTF, we divided individual cells into 4 groups based on the cell order ranks from 1D spatial reconstruction map for scRNA-seq and scATAC-seq data, separately, to satisfy the matching datasets. Then we created 4 "pseudo" bulk replicates, for RNA-seq and ATAC-seq separately, by aggregating cells from the same group. Next, we ran diffTF using the mouse HOCOMOC v10 database with the following parameters: *nPermutations: 100, nBootstraps: 0, nCGBins: 10, RNASeqIntegration: true, pairedEnd: true, peakType: "narrow", minOverlap: 2*.

Transcription Factor Footprint Identification

Footprints are generated by TFs bound to DNA, preventing the Tn5 transposase from cutting DNA in nucleosome-free regions. HINT-ATAC (Li et al. 2019), an HMM-based algorithm, was employed to identify TF binding sites with footprints for HC- and PC/DC-populations. In preparation for running HINT-ATAC, peak files and indexed bam files for HCs and PC/DC were generated by aggregating cells from the same clusters. We started with calling footprints for HCs and PC/DCs separately by running the function *rgt-hint footprinting* with the following parameters: *--atac-seq --paired-end --organism=mm10*. Next, TF-associated footprints for each cluster were identified by detecting motifs from the mouse HOCOMOCO v10 database overlapping with predicted footprints using the *rgt-motifanalysis matching* function. Finally, we generated average scATAC-seq profiles around binding sites of each TF for the two clusters separately by running the *rgt-hint differential* function with the following parameters: *--organism=mm10 --bc --nc 2*. We applied the same analytical workflow and predicted footprints among 4 different states of HCs identified by CellTrails.

Gene Regulatory Network Inference

A gene regulatory network (GRN) is a directed graph, which is made up of nodes (e.g. genes, TFs) with directed links between them. Inferring regulatory relationships between TFs and putative downstream target genes is essential for understanding cell differentiation and development. Up to date, numerous algorithms exist to reconstruct GRNs. However, there are few methods integrating gene expression and chromatin accessibility profiles at single-cell resolution. We followed SCENIC algorithm (Aibar et al. 2017) and developed a 3-step pipeline to reconstruct GRNs. We took HCs and PC/DCs as an example to reconstruct the regulatory landscape.

The first step is to identify co-expression modules from scRNA-seq data using GENIE3 (Huynh-Thu et al. 2010). Briefly, GENIE3 decomposes the network into a lot of regressions. For each regression, the gene expression level of a target gene is predicted from the expression patterns of the TFs using tree-based methods such as Random Forest or Extra-Trees. We applied GENIE3 to identify co-expression modules for HCs and PC/DCs. For running GENIE3, we selected 1,846 genes, including DEGs for both cell types and TF genes identified from the mouse HOCOMOCO v10 database, and generated a gene-by-cell matrix as input for the GENIE3 algorithm. We calculated the Spearman correlation between each pair of genes. Next, we ran GENIE3 to determine co-expression modules and integrated the correlations with the following parameters: *treeMethod = "RF", threshold = 0.005*. The threshold parameter filtered out the

TF/target gene links with a weight below 0.005. In total, GENIE3 yielded 103,041 TF/target gene links for further analysis.

The second step is to identify direct target genes of TFs, which includes setting up a quality control for GENIE3 by scanning putative TF binding sites within accessible regions and removing the links lacking TF binding sites. Specifically, we performed an analysis of TF binding using FIMO (Grant et al. 2011) for HCs and PC/DCs with default settings and the mouse HOCOMOCO v10 database. The TF-target gene links which lacking putative TF binding sites within an upstream and downstream 50kb window of the target gene TSS were filtered out. Next, we defined TF regulons, groups of genes regulated by TFs, by considering all possible interacted genes associated with the respective TFs. Furthermore, we filtered out regulons in which the number of downstream target genes was less than 10 or greater than 900 to keep the methods robust. Based on previous classifications, activator and repressor regulons were determined separately. Additionally, since GENIE3 does not consider autoregulation, we complemented the regulons with autoregulated genes by checking whether the TF binding sites are within in the 100kb window of the TSS of the same TF gene.

The third step is to calculate regulon enrichment scores for each individual cell using AUCell (Aibar et al. 2017). Briefly, AUCell, a ranking-based method, uses the “Area Under the Curve” (AUC) of the recovery curve to determine the enrichment of regulons for individual cells. We first ranked all genes for each cell using the function *AUCell_buildRanking* with default settings. Next, AUC for each regulon in each cell was calculated using *AUCell_calcAUC* function, and only the top 1% of the genes in the ranking were used. Additionally, hierarchical clustering was conducted for both regulons and cells and was added in the heatmap.

We followed the 3-step pipeline to identify TFs controlling IHC and OHC differentiation by predicting the GRN. We selected IHC-specific genes, OHC-specific genes, and TF genes from the mouse HOCOMOCO v10 database. GENIE3 was applied to identify co-expression modules with parameters: *treeMethod = "RF", threshold = 0.001*. In step 2, we performed FIMO motif scanning algorithm for HC peaks only. We filtered out regulons in which the number of downstream target genes was less than 10 or greater than 200. To calculate regulon enrichment scores for IHCs and OHCs, we used exactly the same settings as the previous analysis.

1D Spatial Reconstruction of HCs

We resolved the anatomical positions of HCs at the single-cell level from scRNA-seq and scATAC-seq data by following the conceptual idea previously published for the reconstruction of the mouse organ of Corti from single-cell qPCR data (Waldhaus et al. 2015). First, we identified 427 DEGs ($P < 0.005$) and 127 DARs ($P < 0.001$) between apical and basal compartments from scRNA-seq and scATAC-seq data, respectively. While loosening the threshold would include more DARs, a lot of false positive DARs would likely be included. To deal with the problem, we ran permutation test 100 iterations. Specifically, we permutated the cells 100 times while keeping the original apical/basal cell ratio. For each iteration, peaks were called for permutated apical and basal cells, separately, and DARs were determined using the function *findDAR* from SnapATAC. We used the boxplot to show the number of DARs identified for 100 iterations under different cutoffs.

The hypothesis is that using spatially differential features (e.g. DEGs and DARs) can resolve the anatomical position of individual cells. Next, we projected the individual cells onto a 2D PCA coordinate system and calculated the centroid locations for apical and basal cells, separately. 2D PCA space was rotated based on the apex-to-base centroid-centroid-vector with the apex facing up. Finally, HCs were placed according to their rank order of PC1, indicating each individual cell's relative position along the tonotopic axis for both scRNA-seq and scATAC-sec data.

Prediction of *cis*-regulatory Interactions

We applied Cicero (Pliner et al. 2018) to calculate peak-to-peak coaccessibility from scATAC-seq data for HC and PC/DC cluster, separately. Briefly, Cicero aggregates accessibility profiles for individual cells from the same cluster and calculated a regularized correlation score for each pair of peaks within a 500kb window using a Graphical Lasso model. We started with projecting the cell-by-peak matrix, identified from SnapATAC, into a lower-dimensional space by using the function *reduceDimension* with *num_dim=6*. Next, we ran Cicero for accessible regions to compute coaccessibility scores using the default parameters. We then zoomed in a few genomic loci, related to genes *Sox9*, *S100b*, and *Zbtb18*, and visualized the *cis*-regulatory interactions within these loci in different clusters with a lenient cutoff (*coaccess_cutoff=0*). Additionally, putative TF binding sites scanned from FIMO (Grant et al. 2011) were added to the Cicero map to validate the predicted TF-target gene interactions.

IHC and OHC Identification using scRNA-seq Data

We identified IHC and OHC subpopulations, with respect to the scRNA-seq data, based on 7 previously published marker genes (Waldhaus et al. 2015). The 7 genes included *Fgf10*, *Fgf8*, and *S100a1* as IHC specific genes, and *Slc26a5*, *Fgfr3*, *Cdh1*, and *Ocm* as OHC specific genes. We used these 7 marker genes as features to run UMAP with default parameters. Two subpopulations were identified and differential analysis was conducted between the two subclusters using Wilcoxon sum rank test ($P < 0.01$).

HC Chromatin Accessibility Trajectory Inference

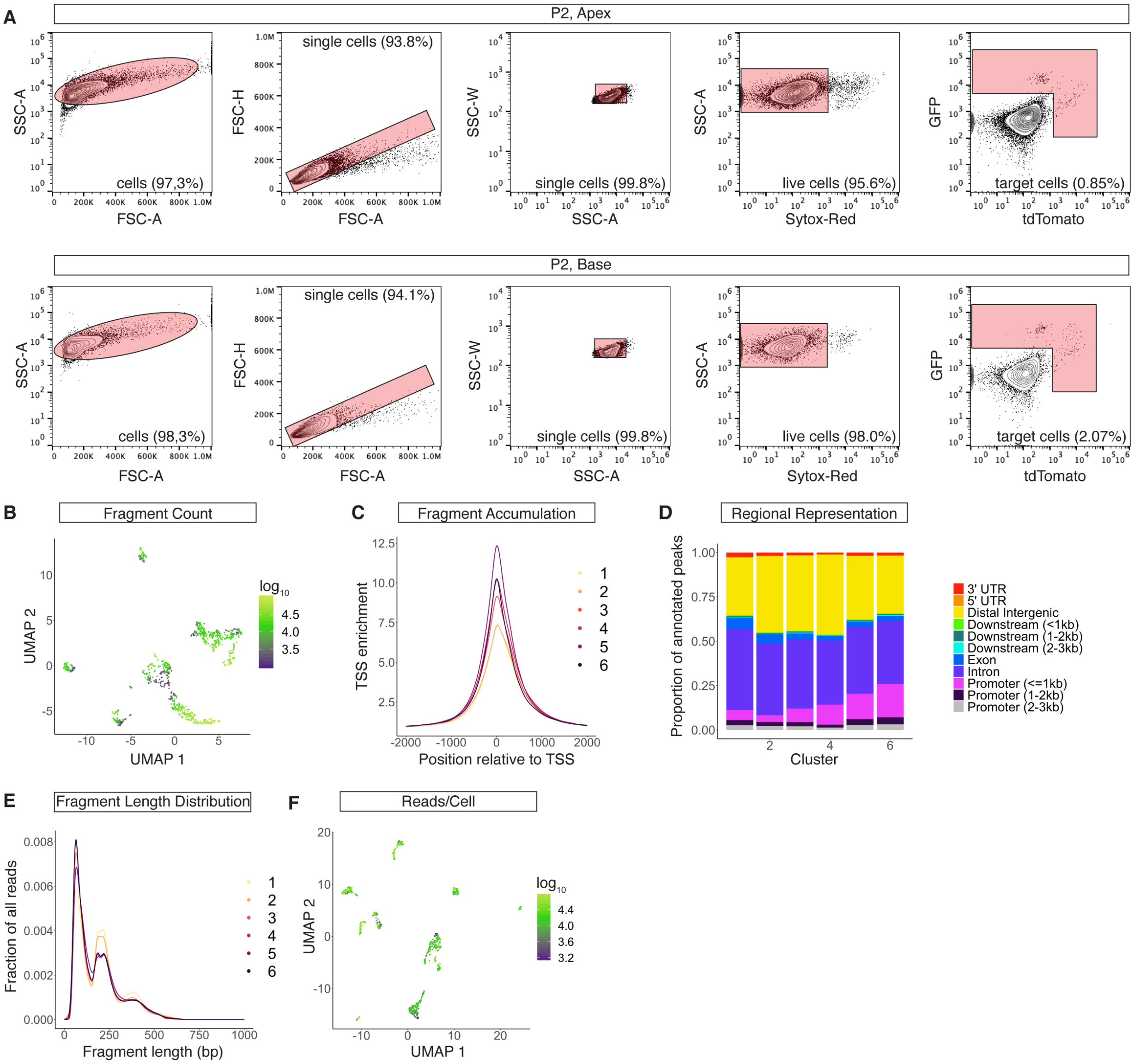
We applied CellTrails (Ellwanger et al. 2018) to reconstruct HC developmental trajectory using *z*-scores from scATAC-seq data. Celltrails employs spectral embedding technique to find a low-dimensional manifold that represents the spatiotemporal relation of cells. To infer the HC differentiation trajectory at epigenomic level, we first prepared a TF-by-cell *z*-score matrix with 426 TF motifs identified from the mouse HOCOMOCO v10 database and 420 HCs. Next, we applied spectral embedding function *embedSample* with default parameters to find a low-dimensional representation and we determined 7 latent variables for further analysis. Hierarchical clustering with a *post-hoc* test was conducted to determine states using the function *findStates* with following parameters: *min_size=0.1, min_feat=5, max_pval=1e-4, min_fc=2*. Then we aligned individual cells to the trajectory using *fitTrajectory* function with default parameters. In summary, we identified 4 states which formed a Y-shape, indicating the bifurcation of HC differentiation. To further investigate the 4 different states, we projected the meta-data information, such as library ID and DAR-based cell rank order, onto the trajectory, and visualized the gradual changes of *z*-scores along the trajectory.

Additionally, we applied Slingshot (Street et al. 2018) to reconstruct and validate trajectory prediction using *z*-scores from scATAC-seq data. Briefly, Slingshot infers the global lineage structure by constructing a cluster-based minimum spanning tree and constructs smooth lineages by fitting similtaneous principle curves. We used the prepared TF-by-cell *z*-score matrix as input for Slingshot and started by conducting dimensionality reduction using PCA and only the first two PCs were subject to further analysis. Next, we identified clusters using Gaussian mixture model from *Mclust* function in the mclust package and three clusters were determined automatically based on the Bayesian information criterion. To reconstruct the trajectory, we run *getLineages* (*start.clus= '1', end.clus=c('2' ,'3')*) function to identify global lineage structure and *getCurves* function with default settings to construct smooth curves and order cells. To validate consistency between Slingshot and CellTrails algorithms, we projected the CellTrails states onto the Slingshot trajectory map. Also, we conducted trajectory analysis using Monocle (Trapnell et al. 2014; Qiu et al. 2017a; Qiu et al. 2017b). Monocle, an unsupervised algorithm, reconstructs the trajectory by constructing minimum spanning tree. Similar to CellTrails and Slingshot, HC TF-by-cell *z*-score matrix was used to run Monocle. We generated a newCellDataSet object with Guassian distribution because the data were already transformed to be normally distributed. Then we conducted dimensionality reduction by running *reduceDimension* function with the following parameters: *max_components = 4,reduction_method = 'DDRTree',norm_method = "none",scaling = T*. Next, we ordered cells using *orderCells* function with default settings. To validate consistency between Monocle and CellTrails algorithms, we projected the CellTrails states onto the Monocle trajectory map.

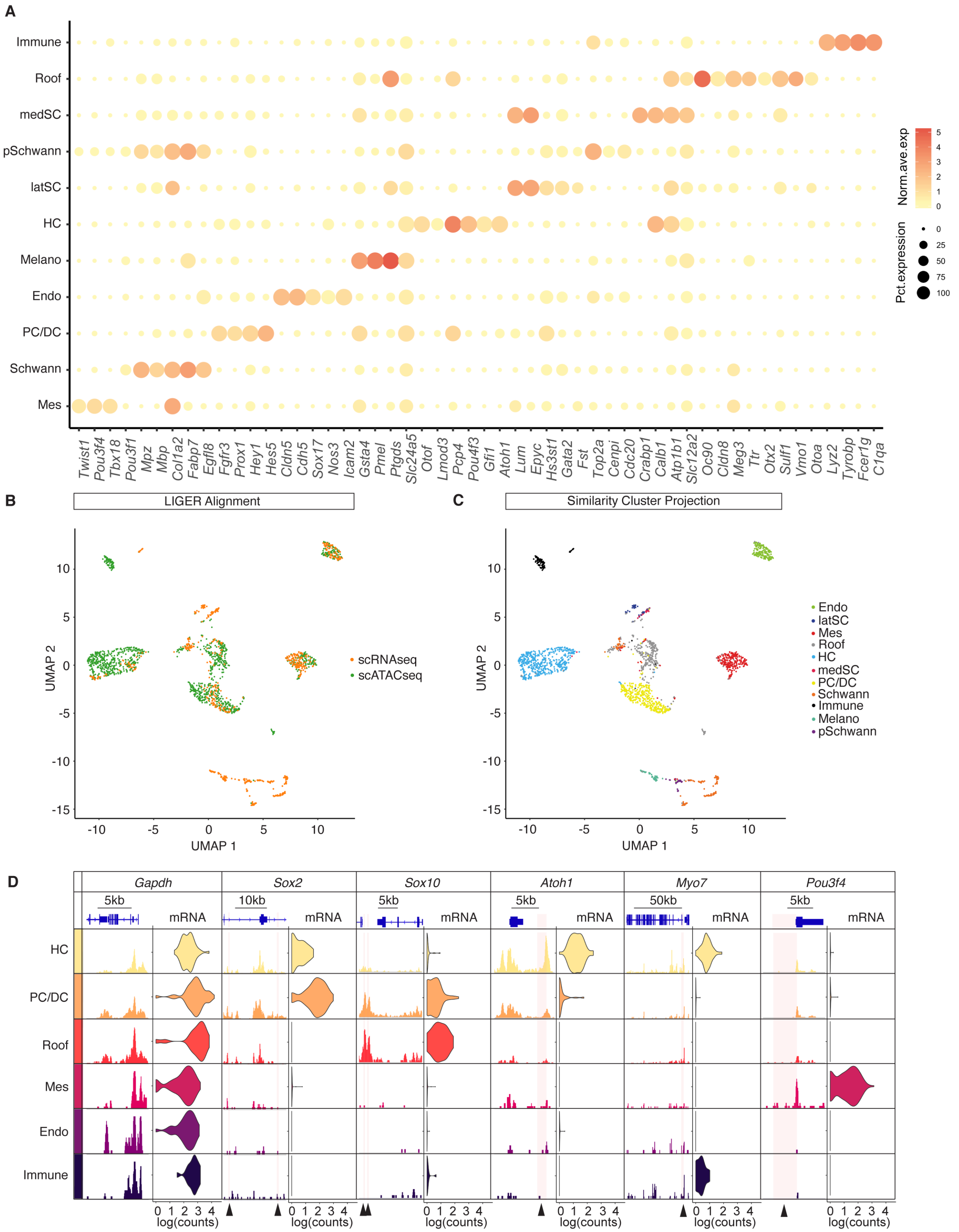
Bulk RNA-seq Analysis

To confirm differential expression for the seven TFs highlighted in the volcano plot (Fig. 6A), we leveraged two previously published bulk RNA-seq data sets. Averaged gene expression values for perinatal HCs (Wiwatpanit et al. 2018) were determined and *P*-values indicated accordingly. Same processing for adult HCs (Li et al. 2018) was performed. To identify differentially expressed genes, we conducted one-side Student *t*-test for each indivudial gene.

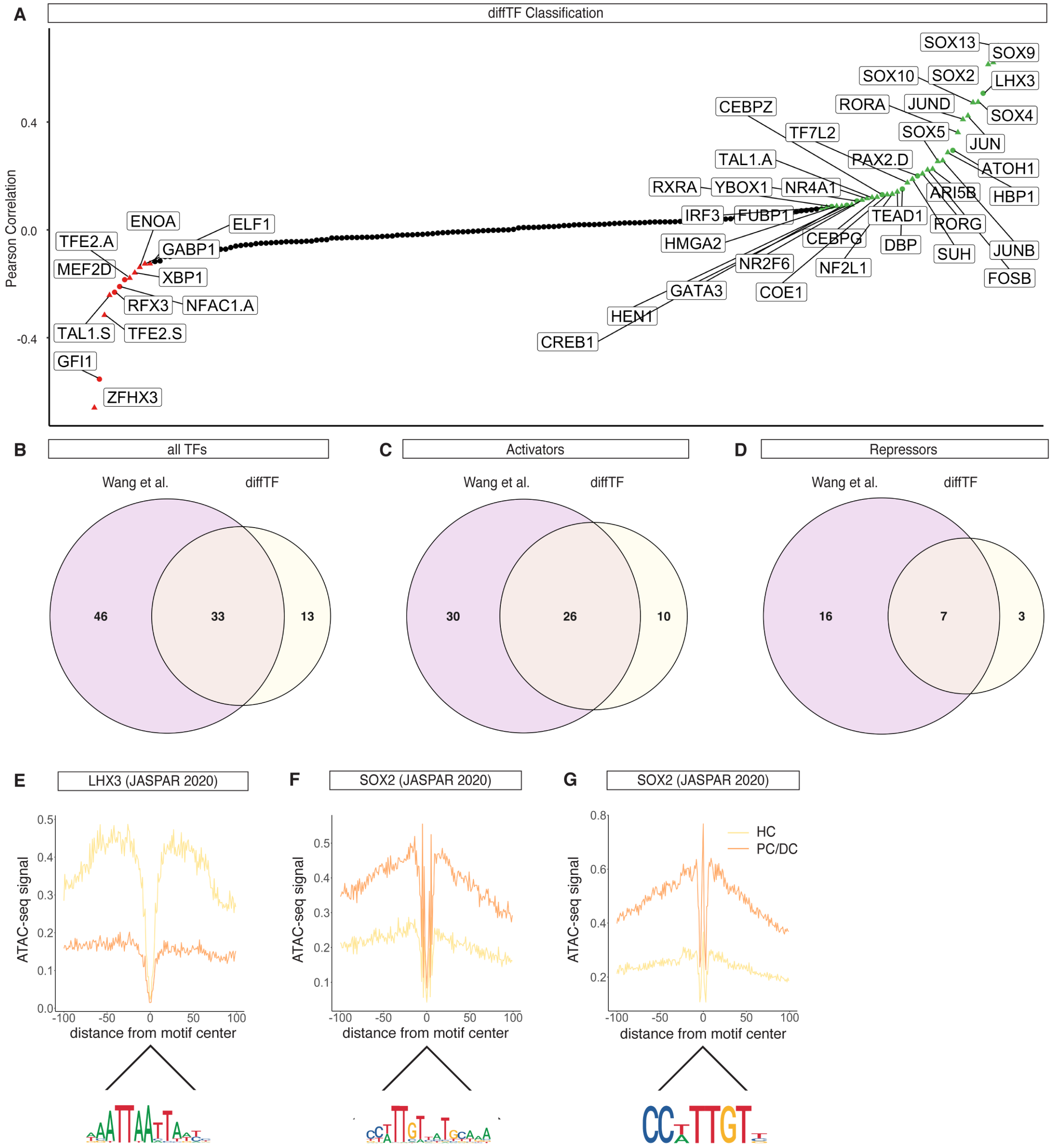
SUPPLEMENTAL FIGURES



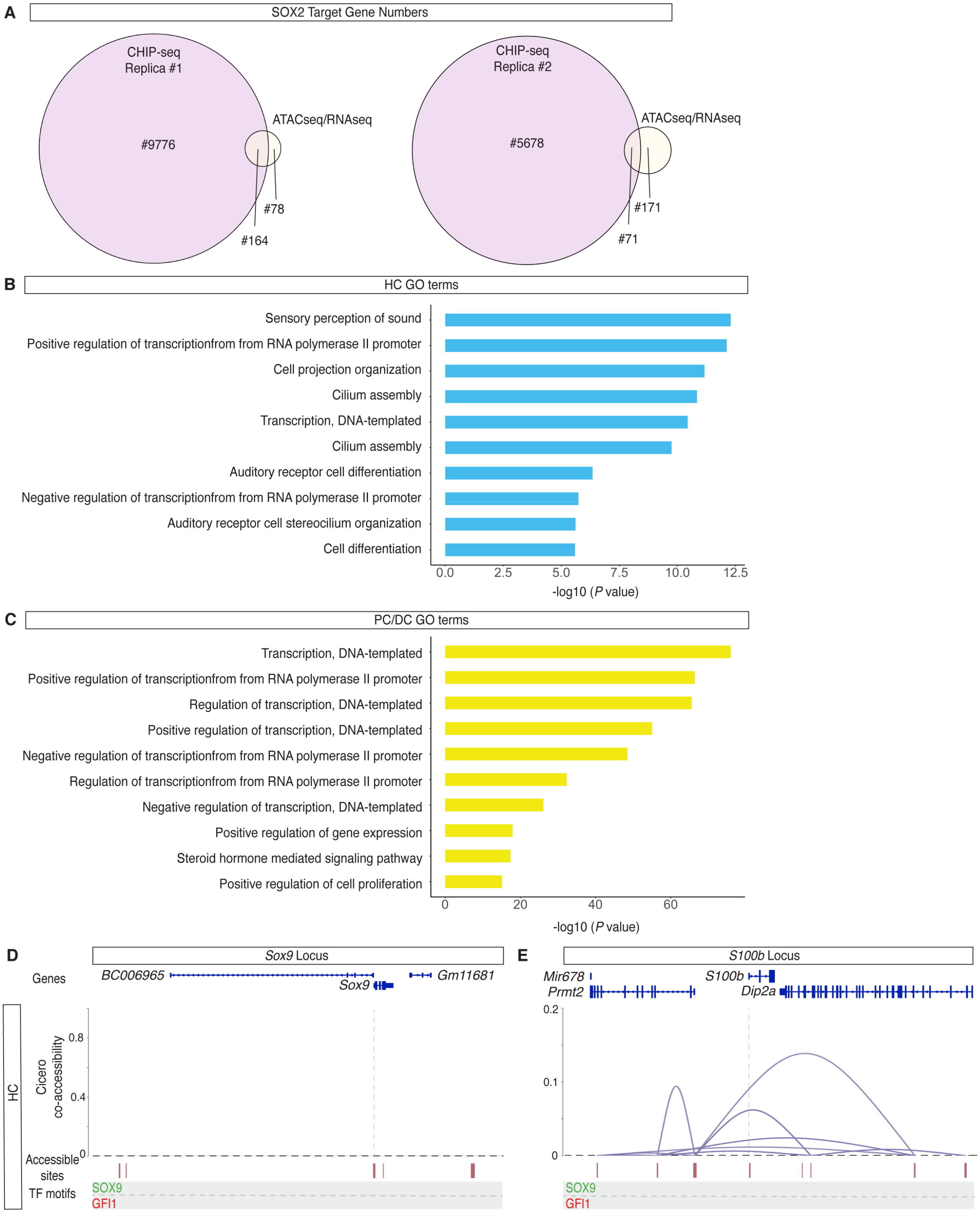
Supplemental Figure S1. Sorting of Apical and Basal Compartment Cells and Quality Control of scRNA-seq and scATAC-seq data. Related to Figure 1. (A) Microdissected cochlear ducts were divided into apical and basal compartments, then processed separately for independent scRNA-seq and scATAC-seq library preparations. Gates were identical between apex and base samples. (B) UMAP plot of unique fragment counts per cell in $\log_{10}$ scale of scATAC-seq dataset. (C) scATAC-seq fragment accumulation relative to TSS position using ataqv. (D) Genomic region composition of DARs for each cluster from scATAC-seq data. (E) Fragment length distribution for each scATAC-seq cluster. (F) UMAP plot of unique read counts per cell in $\log_{10}$ scale of scRNA-seq dataset.



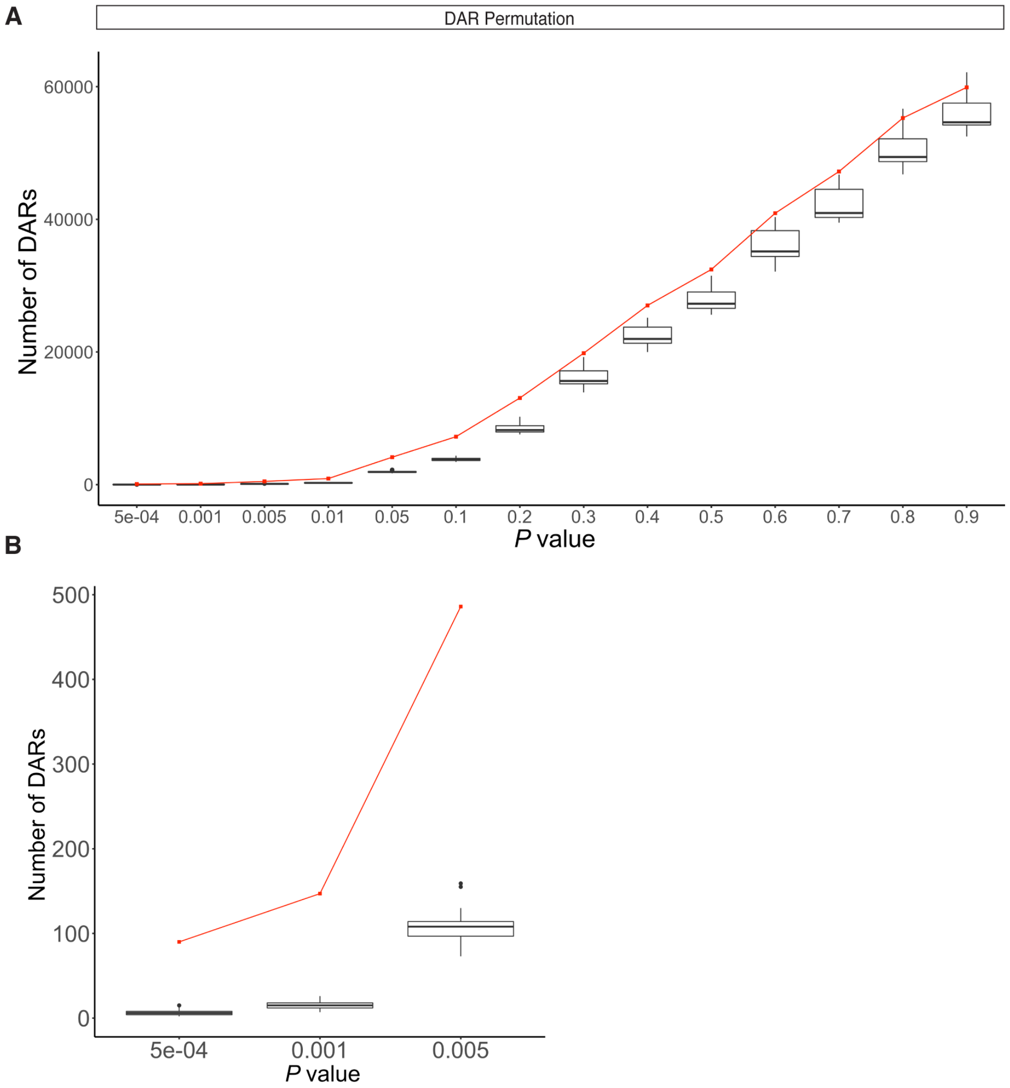
Supplemental Figure S2. scRNA-seq Cell Type Identification and Representation of Previously Published Organ of Corti Regulatory Elements. Related to Figure 1. (A) Dot plot of candidate gene expression levels for each scRNA-seq cluster. The dot size represents percentage of cells expressing a given transcript for the clusters. (B) UMAP plot of LIGER single-cell multi-omics integration analysis, color coded by dataset modality. (C) UMAP plot of LIGER alignment color coded by cell identity from the Jaccard index similarity approach. (D) Enrichment of chromatin accessibility and expression level of candidate genes corresponding to their clusters. Accumulated scATAC-seq fragments at the individual gene locus (left column) and normalized gene expression levels in violin plots (right column) for 6 scATAC-seq clusters. *Gapdh*, a housekeeping gene, revealed a ubiquitous accessibility and expression across 6 clusters. Remaining genes are expressed in organ of Corti cell types. Arrowheads at the bottom of the plot indicate regions of previously published organ of Corti specific regulatory elements (Wilkerson et al. 2019).



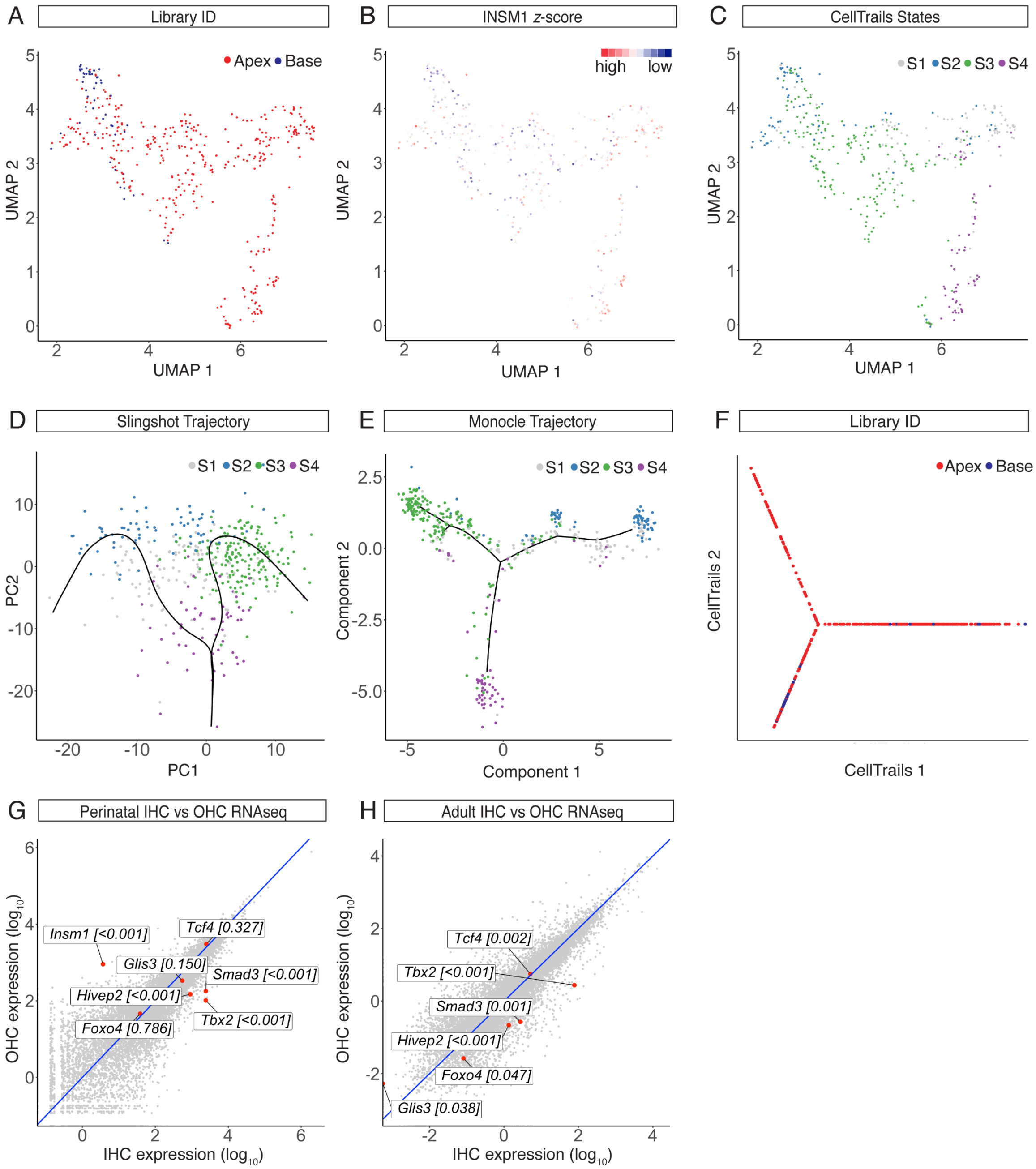
Supplemental Figure S3. diffTF-based Activator/Repressor Classification and Footprint Validation. Related to Figure 2. (A) Transcriptional activators (shown in green) and repressors (shown in red) identified in diffTF algorithm (Berest et al. 2019) by generating 4 "pseudo" bulk ATAC-seq and RNA-seq datasets shown in Pearson correlation. Circles are HC specific TFs, and the triangles are PC/DC specific TFs. (B-D) Venn diagram comparing TF classification presented in Wang *et al.* with diffTF results. (B) All TFs including activators and repressors, (C) activators only, and (D) repressors only. (E-G) TF footprints calculated from scATAC-seq data using HINT-ATAC with JASPAR 2020 consensus sequence. (E) LHX3 footprint. Yellow line represents the HC cluster, and orange line represents the PC/DC cluster. Bottom row: mouse JASPAR 2020 consensus sequence for the LHX3 motif. (F-G) Analogous data representation for two SOX2 motifs. GFI1 and TGIF1 motifs were not curated in JASPAR 2020 and therefore not included in the analysis.



Supplemental Figure S4. HC and PC/DC Regulon Identification and Validation. Related to Figure 3. (A) Venn diagram of quantitative comparison between SOX2 ChIP-seq data (Kwan et al. 2015) and scATAC-seq/scRNA-seq based approach. Shown are numbers of SOX2 downstream target genes identified in two biological replicates from ChIP-seq data and the SOX2 regulon identified in this study. (B) GO term analysis of putative downstream target genes from HC specific regulons using DAVID. (C) GO term analysis of putative downstream target genes from PC/DC specific regulons using DAVID. (D-E) Coaccessibility analysis of *Sox9* and *S100b* loci in the HC cluster using Cicero. (D) In HCs, the peaks associated with *Sox9* TSS are not coaccessible with any of the sites. No putative TF binding sites for SOX9 and GFI1 were annotated based on FIMO motif scanning. Top row: Genome annotation from UCSC Known Genes. Second row: Coaccessibility plot of connections between accessible regions with the TSS. Third row: Accessible regions aligned within the locus. Forth row: putative TF binding sites of SOX9 and GFI1 relative to the accessible regions. TSS position is indicated with a dashed line. (E) Analogous data representation for the *S100b* locus.



Supplemental Figure S5. Permutation Testing of DAR Identification between Apex and Base. Related to Figure 4. (A-B) Number of identified DARs plotted against *P*-value for experimental (red line) vs 100-iteration permutated data (black box plots). (A) *P*-values tested from 0.0005 to 0.9. (B) same data visualized with *P*-values from 0.0005 to 0.005.



Supplemental Figure S6. IHC and OHC Subclusters from scRNA-seq and scATAC-seq Experiments. Related to Figure 5. (A) UMAP plot of HC cluster from scATAC-seq data with projection of library ID. Color code: apex – red, base – blue. (B) UMAP plot of HC cluster from scATAC-seq data with projection of TF INSM1 activity *z*-scores. (C) UMAP plot of CellTrails states S1-S4. (D-E) Trajectory reconstruction of HC cluster based on scATAC-seq *z*-scores using (D) Slingshot and (E) Monocle. CellTrails state ID color coded and projected onto both trajectories. (F) Trajectory reconstruction of HC cluster based on scATAC-seq *z*-scores using CellTrails. Library ID color coded and projected onto the trajectory. Color code: apex – red, base – blue. (G-H) Dot plot of preferentially expressed genes in (G) perinatal IHCs and OHCs (Wiwatpanit et al. 2018) and (H) adult IHCs and OHCs (Li et al. 2018) based on bulk RNA-seq data. Differential OHC expression was confirmed for *Insm1* with temporal restriction around birth. *Tcf4* was confirmed as an OHC-specific gene with differential expression in adult HCs. *Tbx2*, *Smad3*, and *Hivep2* were confirmed differentially expressed in IHC at both time points and *Foxo4* turned into a significantly differentially expressed IHC marker in adult HCs. *Glis3* was the only gene flipping preferential expression from IHCs to OHCs between perinatal and adult samples. The gene expression values are on logarithmic scale. Black arrows highlight the 7 TFs from Fig. 6A including one-side Student *t*-tests *P*-values.

SUPPLEMENTAL REFERENCES

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J et al. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**: 1083-1086.

Alvarado DM, Veile R, Speck J, Warchol M, Lovett M. 2009. Downstream targets of GATA3 in the vestibular sensory organs of the inner ear. *Dev Dyn* **238**: 3093-3102.

Basch ML, Ohyama T, Segil N, Groves AK. 2011. Canonical Notch signaling is not necessary for prosensory induction in the mouse cochlea: insights from a conditional mutant of RBPjkappa. *J Neurosci* **31**: 8046-8058.

Berest I, Arnold C, Reyes-Palomares A, Palla G, Rasmussen KD, Giles H, Bruch PM, Huber W, Dietrich S, Helin K et al. 2019. Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep* **29**: 3147-3159.e3112.

Bioconductor Core Team aPM. 2018. TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s). R package version 3.4.4.

Cai Q, Vethanayagam RR, Yang S, Bard J, Jamison J, Cartwright D, Dong Y, Hu BH. 2014. Molecular profile of cochlear immunity in the resident cells of the organ of Corti. *J Neuroinflammation* **11**: 173.

Carlson M. 2018. org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.7.0.

Cheng C, Wang Y, Guo L, Lu X, Zhu W, Muhammad W, Zhang L, Lu L, Gao J, Tang M et al. 2019. Age-related transcriptome changes in Sox2+ supporting cells in the mouse cochlea. *Stem Cell Res Ther* **10**: 365.

Chung JW, Kang HH, Shin JE, Kim JU. 2004. Accumulation of hypoxia-inducible factor-1alpha in mouse inner ear by noise stimulation. *Neuroreport* **15**: 2353-2356.

Doetzlhofer A, Basch ML, Ohyama T, Gessler M, Groves AK, Segil N. 2009. Hey2 regulation by FGF provides a Notch-independent mechanism for maintaining pillar cell fate in the organ of Corti. *Dev Cell* **16**: 58-69.

Durruthy-Durruthy R, Gottlieb A, Hartman BH, Waldhaus J, Laske RD, Altman R, Heller S. 2014. Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell* **157**: 964-978.

Ebeid M, Huh SH. 2020. Mesenchymal ETV transcription factors regulate cochlear length. *Hear Res* **396**: 108039.

Ellwanger DC, Scheibinger M, Dumont RA, Barr-Gillespie PG, Heller S. 2018. Transcriptional Dynamics of Hair-Bundle Morphogenesis Revealed with CellTrails. *Cell Rep* **23**: 2901-2914.e2913.

Fang R, Preissl S, Hou X, Lucero J, Wang X, Motamedi A, Shiau A, Mukamel E, Zhang Y, Behrens M et al. 2019. Fast and Accurate Clustering of Single Cell Epigenomes Reveals *Cis*-Regulatory Elements in Rare Cell Types. *Biorxiv doi: 101101/615179*.

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D et al. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**: D87-D92.

Gnedeva K, Hudspeth AJ. 2015. SoxC transcription factors are essential for the development of the inner ear. *Proc Natl Acad Sci U S A* **112**: 14066-14071.

Golden EJ, Benito-Gonzalez A, Doetzlhofer A. 2015. The RNA-binding protein LIN28B regulates developmental timing in the mammalian cochlea. *Proc Natl Acad Sci U S A* **112**: E3864-3873.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.

Hertzano R, Dror AA, Montcouquiol M, Ahmed ZM, Ellsworth B, Camper S, Friedman TB, Kelley MW, Avraham KB. 2007. Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system. *Eur J Neurosci* **25**: 999-1005.

Huang dW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Huang M, Sage C, Tang Y, Lee SG, Petrillo M, Hinds PW, Chen ZY. 2011. Overlapping and distinct pRb pathways in the mammalian auditory and vestibular organs. *Cell Cycle* **10**: 337-351.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**: 115-121.

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**: 2204-2207.

Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ. 2013. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res* **41**: D195-202.

Kwan KY, Shen J, Corey DP. 2015. C-MYC transcriptionally amplifies SOX2 target genes to regulate self-renewal in multipotent otic progenitor cells. *Stem Cell Reports* **4**: 47-60.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**: e1003118.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li Y, Liu H, Giffen KP, Chen L, Beisel KW, He DZZ. 2018. Transcriptomes of cochlear inner and outer hair cells from adult mice. *Sci Data* **5**: 180199.

Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. 2019. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**: 45.

Lopez-Juarez A, Lahlou H, Ripoll C, Cazals Y, Brezun JM, Wang Q, Edge A, Zine A. 2019. Engraftment of Human Stem Cell-Derived Otic Progenitors in the Damaged Cochlea. *Mol Ther* **27**: 1101-1113.

Luo XJ, Deng M, Xie X, Huang L, Wang H, Jiang L, Liang G, Hu F, Tieu R, Chen R et al. 2013. GATA3 controls the specification of prosensory domain and neuronal survival in the mouse cochlea. *Hum Mol Genet* **22**: 3609-3623.

Maass JC, Gu R, Cai T, Wan YW, Cantellano SC, Asprer JS, Zhang H, Jen HI, Edlund RK, Liu Z et al. 2016. Transcriptomic Analysis of Mouse Cochlear Supporting Cell Maturation Reveals Large-Scale Changes in Notch Responsiveness Prior to the Onset of Hearing. *PLoS One* **11**: e0167286.

Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR et al. 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**: 133-140.

Maeda Y, Kariya S, Uraguchi K, Takahara J, Fujimoto S, Sugaya A, Nishizaki K. 2020. Immediate changes in transcription factors and synaptic transmission in the cochlea following acoustic trauma: A gene transcriptome study. *Neurosci Res*.

Orchard P, Kyono Y, Hensley J, Kitzman JO, Parker SCJ. 2020. Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. *Cell Syst* **10**: 298-306.e294.

Perl K, Shamir R, Avraham KB. 2018. Computational analysis of mRNA expression profiling in the inner ear reveals candidate transcription factors associated with proliferation, differentiation, and deafness. *Hum Genomics* **12**: 30.

Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A et al. 2018. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**: 858-871.e858.

Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**: 309-315.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017b. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**: 979-982.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Radde-Gallwitz K, Pan L, Gan L, Lin X, Segil N, Chen P. 2004. Expression of Islet1 marks the sensory and neuronal lineages in the mammalian inner ear. *J Comp Neurol* **477**: 412-421.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24-26.

Rose MF, Ren J, Ahmad KA, Chao HT, Klisch TJ, Flora A, Greer JJ, Zoghbi HY. 2009. Math1 is essential for the development of hindbrain neurons critical for perinatal breathing. *Neuron* **64**: 341-354.

Sanz C, León Y, Cañón S, Alvarez L, Giraldez F, Varela-Nieto I. 1999. Pattern of expression of the jun family of transcription factors during the early development of the inner ear: implications in apoptosis. *J Cell Sci* **112 (Pt 22)**: 3967-3974.

Scheffer DI, Shen J, Corey DP, Chen ZY. 2015. Gene Expression by Mouse Inner Ear Hair Cells during Development. *J Neurosci* **35**: 6366-6380.

Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. 2017. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**: 975-978.

Simonoska R, Stenberg AE, Duan M, Yakimchuk K, Fridberger A, Sahlin L, Gustafsson JA, Hultcrantz M. 2009. Inner ear pathology and loss of hearing in estrogen receptor-beta deficient mice. *J Endocrinol* **201**: 397-406.

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**: 477.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**: 1888-1902.e1821.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**: 381-386.

Waldhaus J, Durruthy-Durruthy R, Heller S. 2015. Quantitative High-Resolution Cellular Map of the Organ of Corti. *Cell Rep* **11**: 1385-1399.

Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. 2019. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**: 1873-1887.e1817.

Wilkerson BA, Chitsazan AD, VandenBosch LS, Wilken MS, Reh TA, Bermingham-McDonogh O. 2019. Open chromatin dynamics in prosensory cells of the embryonic mouse cochlea. *Sci Rep* **9**: 9060.

Wiwatpanit T, Lorenzen SM, Cantú JA, Foo CZ, Hogan AK, Márquez F, Clancy JC, Schipma MJ, Cheatham MA, Duggan A et al. 2018. Trans-differentiation of outer hair cells into inner hair cells in the absence of INSM1. *Nature* **563**: 691-695.

Yang S, Cai Q, Bard J, Jamison J, Wang J, Yang W, Hu BH. 2015. Variation analysis of transcriptome changes reveals cochlear genes and their associated functions in cochlear susceptibility to acoustic overstimulation. *Hear Res* **330**: 78-89.

Young KM, Mitsumori T, Pringle N, Grist M, Kessaris N, Richardson WD. 2010. An Fgfr3-iCreER(T2) transgenic mouse line for studies of neural stem cells and astrocytes. *Glia* **58**: 943-953.

Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382-2383.

Zhang Y, Chen D, Zhao L, Li W, Ni Y, Chen Y, Li H. 2019. Nfatc4 Deficiency Attenuates Ototoxicity by Suppressing Tnf-Mediated Hair Cell Apoptosis in the Mouse Cochlea. *Front Immunol* **10**: 1660.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.