PNAS
www.pnas.org

Supplementary Information for

POU4F3 Pioneer Activity Enables ATOH1 to Drive Diverse Mechanoreceptor Differentiation Through a Feed-Forward Epigenetic Mechanism

Haoze V. Yu[1, *], Litao Tao[1, *], Juan Llamas[1], Xizi Wang[1], John D. Nguyen[1], Talon Trecek[1], Neil Segil[1, 2, *]

[1]Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine of the University of Southern California, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Biology at USC, CA 90033, USA.

[2]Caruso Department of Otolaryngology – Head and Neck Surgery, Keck School of Medicine of the University of Southern California, CA 90033, USA.

*These authors contributed equally

[†]**Corresponding author:** Neil Segil

Email: nsegil@med.usc.edu

This PDF file includes:

Supplementary Materials and Methods

Figures S1 to S10

Legends for Datasets S1 to S3

Supplementary reference list

Supplemental Materials and Methods

Animals

The experimental procedures were approved by the Animal Care and Use Committee of the University of Southern California. *Cdkn1b-Gfp* BAC transgenic (CD1 background) (26), *Atoh1-Gfp* knock-in (C57BL/6 background) (27), *Pou4f3-Dtr* knock-in (FVB/NJ background) (79), *Atoh1-flox* knock-in (C57BL/6 background) (80), and *Pax2-Cre* BAC transgenic (CD1 background) (45) mouse lines were used in this study. Expression of POU4F3 was disrupted in $Pou4f3^{Dtr/Dtr}$ animals due to the insertion of *hDTR* gene at the start codon of *Pou4f3* gene (79). POU homozygous mutant mice used in this study were also $Atoh1^{Gfp/Gfp}$. F-actin enriched stereocilia of the utricular hair cells was stained by Phalloidin-iFluor (Abcam, ab176759; diluted 1/1,000 in Phosphate Buffered Saline (PBS) containing 0.05% Triton X-100), and the absence of stereocilia staining in $Pou4f3^{Dtr/Dtr}$ utricles allowed fast identification of *Pou4f3* homozygous mutant, from heterozygous mutant and wild type littermates (36, 37). Conditional *Atoh1* knockout embryos were generated by crossing $Atoh1^{flox/+}$; $Pax2^{Cre/+}$ animals to $Atoh1^{flox/flox}$ animals.

Fluorescence-activated cell sorting

Cochlear tissue was incubated with trypsin (Invitrogen, 0.25%) at 37°C for 8 minutes. To stop the reaction, Fetal Bovine Serum (FBS) (Invitrogen) at a final concentration of 2% was added. The tissues were dissociated into single-cell suspension by trituration with a P200 pipette for 3 min, and then filtered with a 40μm cell strainer (VWR International). DAPI (1:10,000; Invitrogen) was added to the single-cell suspension to mark dead cells during FACS purification.

To purify Merkel cells and their progenitors, skin tissue (head, limbs and tail removed) from E17.5 mouse embryos was digested with 5mg/mL dispase II (Sigma, D4693) in PBS at 37 °C for 45 minutes. The epidermal tissues were separated from dermis with forceps, and then incubated with trypsin (Invitrogen, 0.25%) at 37 °C for 10 minutes. To stop the reaction, FBS (Invitrogen) was added to a final concentration of 5%. The tissue was dissociated into a single-cell suspension by trituration with P1000 pipette, and then passed through a 40μm cell strainer (VWR International). DAPI was added to mark dead cells during FACS purification.

Merkel cells were purified by their *Atoh1*-GFP expression. Touch dome epithelial progenitors were FACS-purified by elimination of Merkel cells (GFP+), and staining for progenitors with antibodies to Integrin-α6, Cd34, Sca-1, and CD200 as described previously (54).

Fluorescence-activated cell sorting was done on an Aria II (BD Biosciences). Cells were collected either into PBS with 5% FBS for μATACseq and C&R, or into RNA lysis buffer (Zymo Research, R1060) for RNAseq. The purity of the target population was confirmed to be >90% by re-sort, and/or cell counting.

μATACseq

μATACseq is based on ATACseq (28), with some modifications, including the elimination of the nuclear purification step. Hyperactive Tn5 was produced as previously described (81). The transposition activity of homemade Tn5 was assessed, and 1 unit of activity was defined as the amount of transposase that was able to convert 100 ng λDNA to fragments of an average size of 200-400 bp in 15 minutes at 55˚C. FACS-purified cells (500 to 3000 cells per reaction) were centrifuged to remove the supernatant. After pelleting cells at 600G for 5 min, 20 μL of lysis buffer (10 mM Tris HCl (pH 7.4), 5 mM $MgCl_2$, 10% DMF, 0.2% N-P40) was added to resuspend the pellet, followed by pipetting 6-10 times to release the nuclei without purification, and 30 μL reaction buffer (10 mM Tris HCl (pH 7.4), 5 mM $MgCl_2$, 10% N,N-dimethylformamide, plus 1 unit of transposase) was mixed on a Vortex for 5 seconds. The reaction was incubated at 37 °C for 20 minutes, followed by addition of 350 μL Enzymatic Reaction Cleanup buffer (MiniElute, Qiagen) to stop the reaction. After DNA fragment purification (MiniElute, Qiagen), ATACseq libraries were constructed as previously described (28).

CUT&RUN (C&R)

Genome-wide histone modifications and transcription factor bindings were assayed by the CUT&RUN technique according to the original protocol (32). Approximately 5,000 cells were used for histone modifications, and ~10,000 cells for transcription factor binding studies. The following antibodies were used: rabbit-anti-H3K4me1 (Active Motif, #61633), rabbit-anti-H3K27ac (Active motif, #39133), rabbit-anti-H3K27me3 (Active Motif, #39155), rabbit-anti-H3K4me3 (Active Motif, #39159), rabbit-anti-GFP (Torrey Pines Biolabs, TP401), mouse-anti-POU4F3 (Santa Cruz, sc-81980), and secondary rabbit-anti-mouse antibody (Abcam, ab46540).

DNA fragments extracted from CUT&RUN samples were made into sequencing libraries using Accel-2S DNA library prep kit (Swift Biosciences) with Molecular Identifier (MID) and single index adapters. DNA concentration and fragment distribution were measured on a Tapestation (Agilent), using high-sensitivity D5000 tape. Libraries were pooled to equal molarities and then diluted to 4nM, for sequencing.

Analysis of μATACseq and CUT&RUN data

Libraries from μATACseq and C&R were sequenced using the NextSeq 500 platform (Illumina) and a minimum of 20 million paired-reads/sample was collected. The Encode analysis pipelines (https://github.com/ENCODE-DCC/chip-seq-pipeline) for ATACseq and ChIPseq data, were modified and used to analyze our sequencing data. Briefly, the raw reads were trimmed to 37bp and aligned to GRCm38/mm10 genome assembly by Spliced Transcripts Alignment to a Reference (STAR) aligner (82). PCR duplicates were removed, based on genomic coordinates for ATACseq, or by MIDs using UMI-tools for C&R; reads that aligned to “blacklist regions” (83) were

removed, and then peaks were called by Model-based analysis of ChIP-Seq (MACS2) (84) with p=0.01 cutoff and disabled dynamic lambda option (--nolambda) for individual replicates. Peaks from individual replicates were further filtered by IDR<0.01, and the overlapping peaks between replicates were used. Bigwig files were generated from dup-removed BAM files with *bedtools* and *bedgraphtobigwig*, and normalization was based on total read numbers. Heatmaps and their quantitative average signal profiles were generated with DeepTools (85) based on the normalized bigwig signal files. Individual genomic loci were examined by IGV (Broad Institute). HOMER (86) was used to identify *de novo* motifs. Differential analyses of ATACseq data was done with DESeq2 (87). Gene ontology analyses were done by EnrichR (88) and GREAT (89).

RNAseq

Total RNA was extracted from FACS-purified cells with Quick-RNA Microprep kit (Zymo Research), quantified on a Tapestation (Agilent), and processed for library construction with QIAseq FX Single Cell RNA Library Kit (Qiagen). At least 2 biological replicates were collected for each sample/ condition, and over 20 million reads were sequenced for each replicate. Reads were mapped to GRCm38/mm10 genome assembly by STAR aligner (82), and counted against transcript by htseq-count (90). Differentially expressed protein coding genes were identified using DESeq2 (87). FPKM was calculated by Cufflinks (91).

***In vitro* overexpression of transcription factors in MEF**

Mouse embryonic fibroblasts were obtained from E13.5 wild type embryos after carefully removing the organs and the spinal cord. Tissue was minced and enzymatically treated with 0.25% trypsin-Ethylenediaminetetraacetic acid for 30 minutes at 37°C. Trypsin treatment was stopped by 2% FBS (Invitrogen), and tissue was then dissociated by trituration with a P1000 pipette for 3 minutes. The dissociated cells were centrifuged (800g for 10 minutes) and the pellet was suspended in MEF media (Dulbecco's Modified Eagle Medium/F12 and 10% FBS) before plating in gelatin coated T75 tissue culture flasks. The MEFs were expanded in culture and then cryopreserved in liquid nitrogen using freezing media (MEF media plus 10% Dimethylsulfoxide (DMSO)). MEFs were cultured to reach 70-80%% confluency before lentivirus infection. All cell cultures tested negative for mycoplasma contamination.

Virus was produced by transfection of HEK293 with 3 constructs, pPAX2, pVSVG, plus pHR plasmids containing the coding regions for alternately, ATOH1-t2a-GFP, POU4F3-p2a-tdTomato or GFP control (sequences upon request) using FUGENE HD (Promega). Three days after transfection, cells and media were harvested, filtered (0.45µm filter, Sigma). The supernatants containing the lentiviruses were used without further concentration steps.

Organotypic Utricular Culture

Utricles were dissected from postnatal day 2 pups (*Atoh1*$^{flox/flox}$;*Pax2*$^{Cre/+}$ mice) in pre-chilled PBS. The utricular sensory epithelia in *Atoh1* conditionally homozygous-mutant mice were identified morphologically by their sparse innervation at this stage (9, 92), compared to wild type. The utricles from *Atoh1* mutant animals were mounted on SPI black membranes (SPI Supplies) floating on culture medium (DMEM/F12, supplemented with 100U penicillin, 1:100 N2 (Invitrogen), 1:50 B27 (Invitrogen), 5 ng/ mL Epidermal Growth Factor (Sigma) and 2.5 ng/ mL Fibroblast Growth Factor (NIH)), and cultured at 37 °C, 5% CO_2 and 5% O_2, as described previously (78).

AAV used for utricular infection was produced as previously described (46). Three plasmid constructs were transfected into HEK293 cells using FUGENE HD (Promega). These plasmids were dF6 adenoviral vector, pAAV2/ Anc80L65 encoding AAV2 rep and Anc80L65 cap genes, and a plasmid containing the inverted terminal repeat sequences flanking the coding regions for ATOH1-t2a-GFP or GFP control. Three days after transfection, cells and media were harvested and filtered. The AAV virus in the supernatant was concentrated using AAVanced Concentration reagent (SBI, AAV100A). Concentrated AAV-ATOH1-t2a-GFP, or AAV-GFP control virus was added to the culture media. Tissues were fixed 72 hours post-infection for immunohistochemistry.

Immunohistochemistry

Epidermis preparation: Skin was prepared from E17.5 mouse embryo trunk regions (head, limbs and tail removed) and treated with 5mg/mL dispase II (Sigma, D4693) in PBS at room temperature for 20 minutes. The epidermis was carefully separated from dermis using forceps, then cut into small pieces. Tissue was fixed with 4% paraformaldehyde in PBS at room temperature for 15 minutes. Subsequently, the tissues were blocked and permeabilized overnight at 4 °C with 10% donkey serum (Millipore, S30-100ML) in PBS, supplemented with 0.5% Triton X-100 (Sigma, T8787). Primary and secondary antibody incubations were performed sequentially overnight at 4 °C in 1% donkey serum in PBS, supplemented with 0.05% Triton X-100. Washing with PBS at room temperature for 15 min was performed twice after each antibody incubation. The epidermal tissue was mounted on slides with the basal layer facing up.

Cochleae preparation: from E13.5-E15.5 embryos were collected in pre-chilled PBS after dissection. The organs were incubated in PBS containing 1mg/mL dispase II (Sigma, D4693) and 1mg/mL collagenase (Worthington) at 37 °C for 8 minutes, and then cochlear ducts were freed from surrounding mesenchyme using forceps, as previously described (78). Cochleae from E17.5 or older embryo/pups were dissected from inner ears in pre-chilled PBS. The spiral ganglia, Reissner's membrane, and the lateral wall were carefully removed to obtain cochlear surface preps. Cochlear ducts or surface preps were fixed and stained with the aforementioned protocol.

Primary antibodies were diluted at 1:500, and secondary antibodies were diluted at 1:5,000. Primary antibodies used in this study were mouse anti-p27^{kip1} (Thermo Fishers, AHZ0452), rabbit

anti-MYO7A (Proteus BioSciences, 25-6790), chicken anti-GFP (GeneTex, GTX13970), mouse anti-POU4F3 (Santa Cruz, sc-81980), goat anti-SOX2 (Santa Cruz, sc-17320, discontinued), mouse anti-KRT20 (Dako, M701929-2), rabbit anti-Caspase3 (R&D, AF835), mouse anti-LHX3 (DSHB, 67.4E12), rabbit anti-SIX1 (Cell Signaling, 12891s). Secondary antibodies used in this study were from Thermo Fishers or Jackson ImmunoResearch.

Supplemental Figures

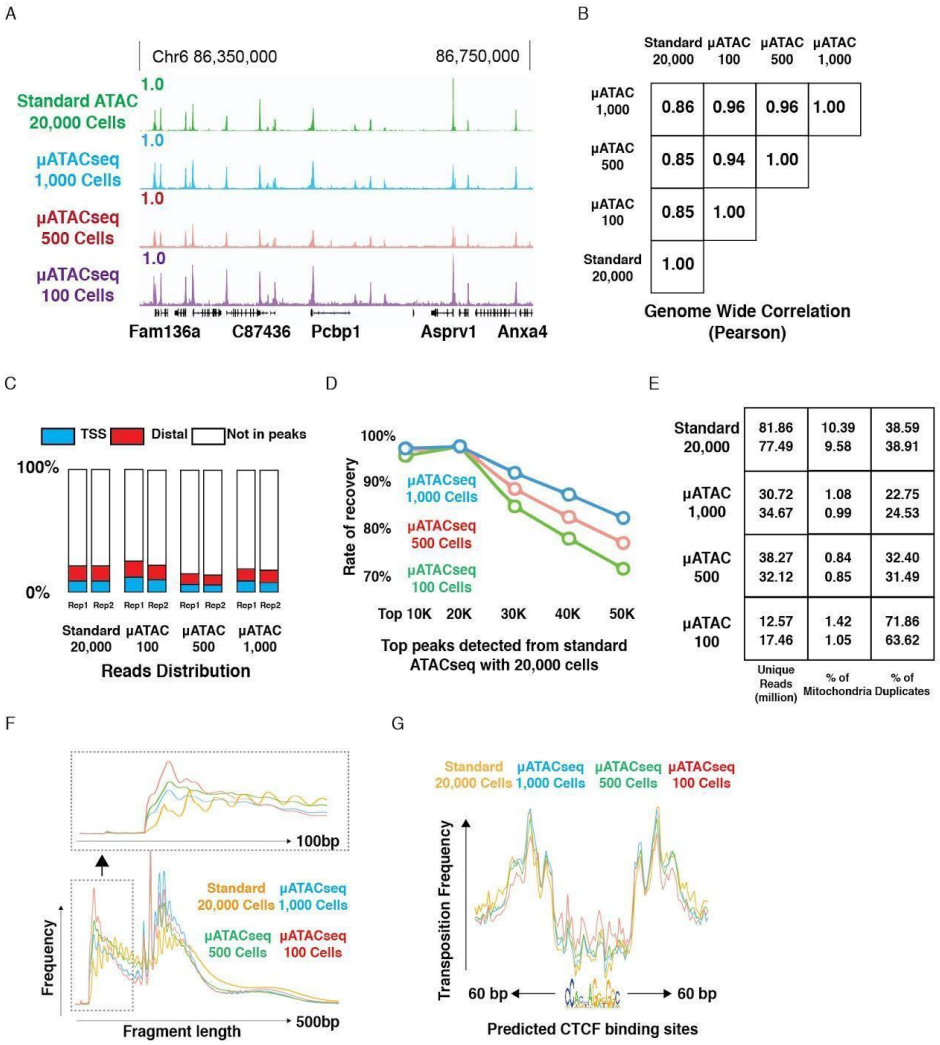

Figure S1. Using *μ*ATACseq to profile chromatin accessibility with as few as 100 cells.
(A) Genome browser representation of the comparison of data collected from the standard ATACseq and *μ*ATACseq protocols with different cell numbers. Data collected from 100 cells using μATACseq compare favorably to those from 20,000 cells with the standard ATACseq protocol.
(B) Pearson correlation among the data collected from the standard ATACseq and μATACseq with different cell numbers.
(C) Histogram representation of read-distribution among data sets collected from the standard ATACseq and *μ*ATACseq (100, 500, and 1,000 cells), shows similar proportions of reads falling in promoters, enhancers and non-peak regions between samples with different cell numbers.
(D) Graph shows percentage of standard ATACseq peaks that are detected by *μ*ATACseq with different numbers of cells in each group. In the 100-cell group, more than 70% of the top 50,000 peaks present in the standard ATACseq data set with 20,000 cells, were detected.
(E) Comparison of the number of uniquely mapped reads, percentage of mitochondrial reads, and the percentage of duplicated reads, of data from the standard ATACseq and *μ*ATACseq experiments with different number of cells.
(F) Comparison of the read-length distribution among the data from the standard ATACseq and *μ*ATACseq experiments with different numbers of cells. On average, the reads are shorter as the cell

number decreases, because fewer cells mean higher enzyme-to-cell ratios, which leads to more complete digestion and shorter reads. However, no over-digestion was observed, as seen by similar mono- and di-nucleosomal fragment distribution in the different samples.
(G) Comparison of the average footprint depth (msCentipede) at the predicted CTCF binding sites among ATACseq (20,000 cells) and *μ*ATACseq in different samples (100, 500, and 1,000 cells). In the 100-cell group, we observed a slightly shallower average footprint at CTCF binding sites, compared to the other groups, consistent with the higher enzyme-to-cell ratio. Nevertheless, the footprints from all four groups were clear and informative, indicating that our protocol is robust with different enzyme-to-cell ratios without causing over-digestion.

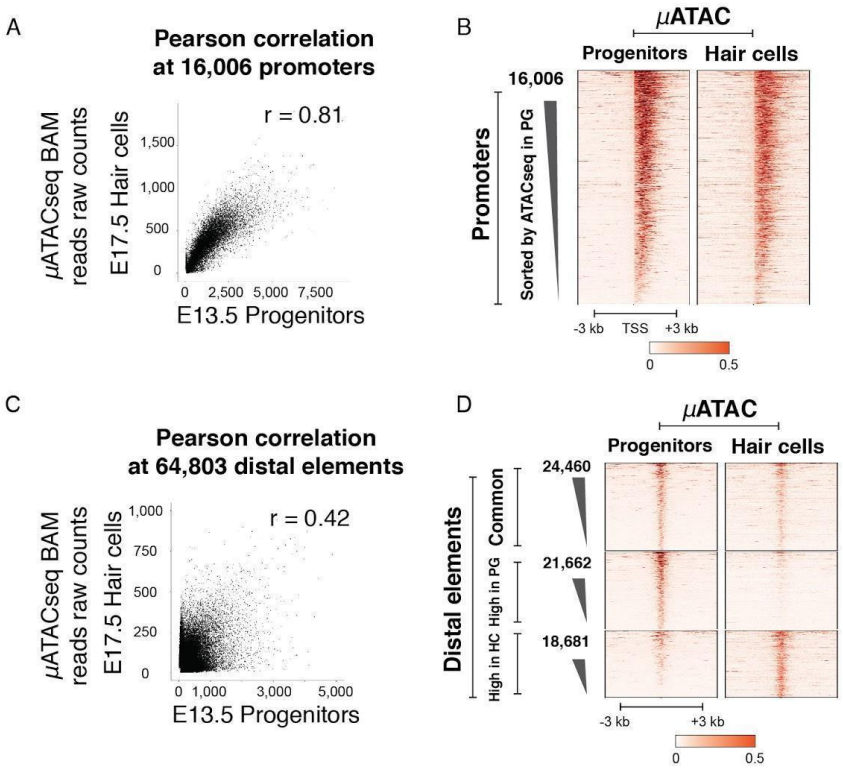

Figure S2. Correlation analysis at the promoters and distal elements detected in the sensory progenitors and hair cells.
(A) Scatter plot shows that the *μ*ATACseq data at the promoter regions are highly correlated between sensory progenitors and hair cells (r=0.81).
(B) Heatmap representation of the *μ*ATACseq results at the promoter regions in sensory progenitors and hair cells. Similar promoter accessibility is observed between hair cells and progenitors.
(C) Scatter plot at the distal elements shows the Pearson correlation of the *μ*ATACseq data between sensory progenitors and hair cells (r=0.42).
(D) Heatmap representation of *μ*ATACseq results at the distal elements in sensory progenitors and hair cells. The three groups of distal elements are defined in Figure 1B.

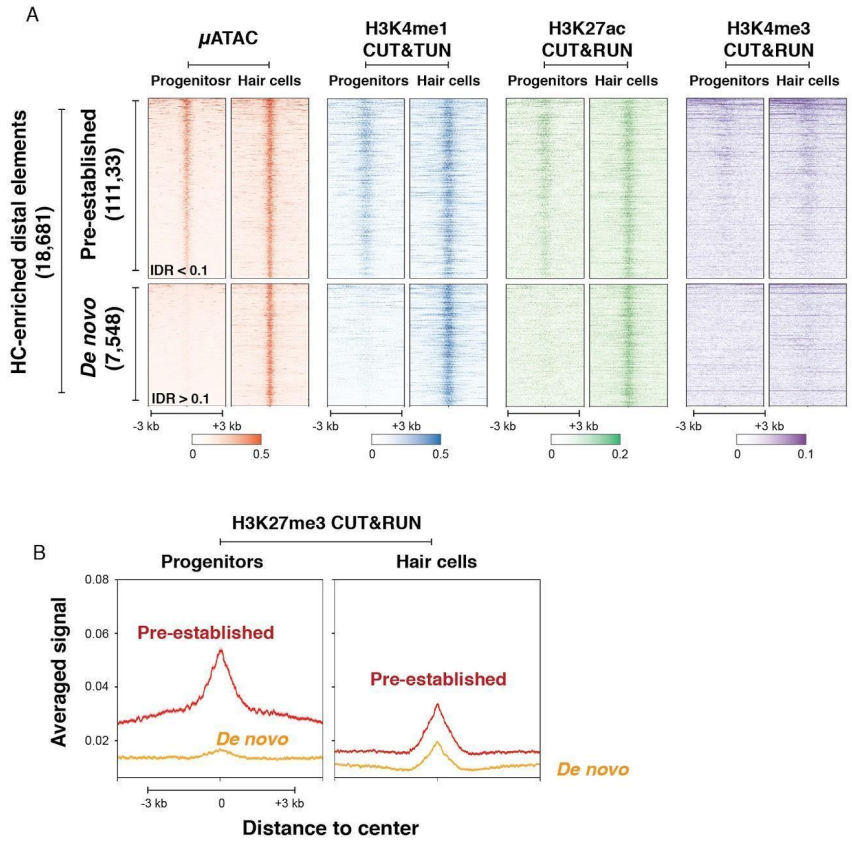

Figure S3. Nucleosomes at the *de novo* hair cell-enriched distal elements are not modified by active or inhibitory marks in progenitors (E13.5), in contrast to pre-established regions.
(A) Heatmap representation of the *μ*ATACseq and CUT&RUN results for active histone marks (H3K4me1, H3K27ac, and H3K4me3) at the hair cell-enriched distal elements.
(B) Signal profile of CUT&RUN results shows the presence of the inhibitory histone mark H3K27me3 at pre-established, but not at *de novo* hair cell-enriched distal elements.

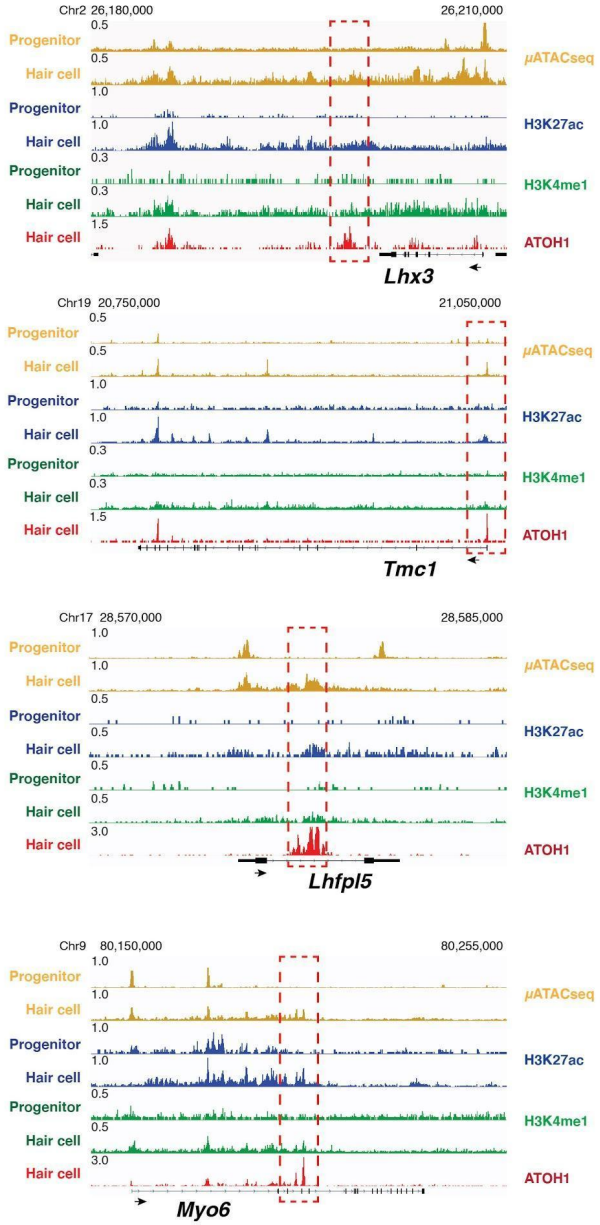

Figure S4. Chromatin accessibility, histone modifications and ATOH1 binding signal at additional gene loci associated with the *de novo* ATOH1 targets in hair cells.
IGV tracks show chromatin accessibility (*μ*ATACseq), H3K27ac, H3K4me1 and ATOH1 CUT&RUN profiles at *Lhx3*, *Tmc1*, *Lhfpl5* and *Myo6* loci. Putative enhancers are indicated by boxes.

A

Expression level of POU domain transcription factors (FPKM)

	PG1	PG2	HC1	HC2	KO1	KO2
Pou1f1	0	0	0	0	0	0
Pou2f1	2.2	1.9	12.5	13.3	7.6	9.1
Pou2f2	0	0	0	0	0	0
Pou2f3	0	0	0	0	0.1	0.2
Pou3f1	0	0	1.0	1.0	11.1	3.8
Pou3f2	2.2	1.8	0.6	0.4	7.3	7.4
Pou3f3	30.2	25.1	29.6	27.2	16.4	12.9
Pou3f4	0	0	2.2	2.4	1.0	10.4
Pou4f1	0	0	1.3	1.8	41.9	19.5
Pou4f2	0	0	0	0	0	0
Pou4f3	0	1.0	432.0	493.3	563.5	576.4
Pou5f1	0	0	0	0	0.3	0
Pou5f2	6.4	3.5	2.8	2.7	7.7	2.3
Pou6f1	2.8	1.4	8.1	7.8	29.1	19.7
Pou6f2	0.3	0	0	0.1	0.6	0.7

B

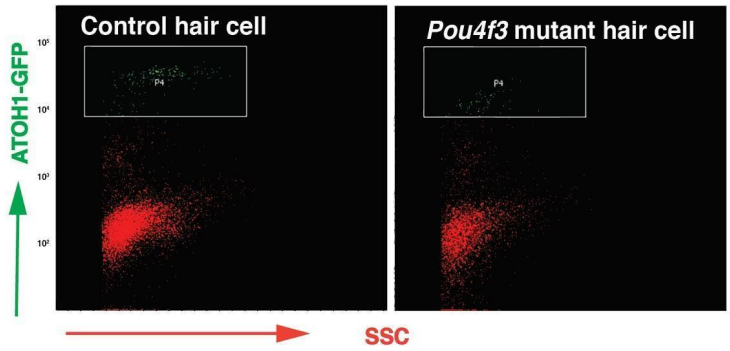

Figure S5. Purification of the *Pou4f3* mutant hair cells using ATOH1-GFP mouse model.
(A) Summary of expression levels (RNAseq) of different POU-domain transcription factors in sensory progenitors, wild type hair cells, and *Pou4f3* mutant hair cells. *Pou4f3* mRNA (green) is detected, but not translated in *Pou4f3*-DTR mutant (KO) hair cells (79). PG, progenitors; HC, Hair cell.
(B) Fluorescence Activated Cell Sorting of ATOH1-GFP positive hair cells (green) from wild type and *Pou4f3* mutant mouse cochlea (see Methods).

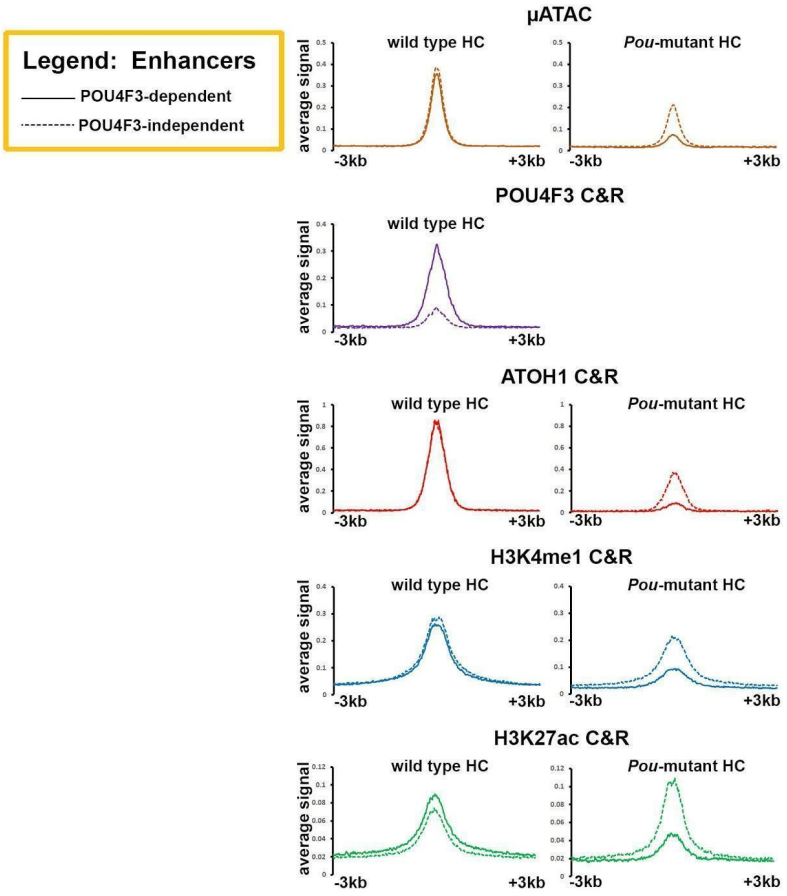

Figure S6. Quantitative signal profiles of heatmaps showing POU4F3-dependent and -independent *de novo* ATOH1 targets from Figure 2.
Average signal profiles of chromatin accessibility, POU4F3 C&R, ATOH1 C&R, H3K4me1 C&R and H3K27ac C&R at the 1,458 POU4F3-dependent and 1,806 POU4F3-independent *de novo* ATOH1 targets, from wild type hair cells and *Pou4f3*-mutant hair cells.

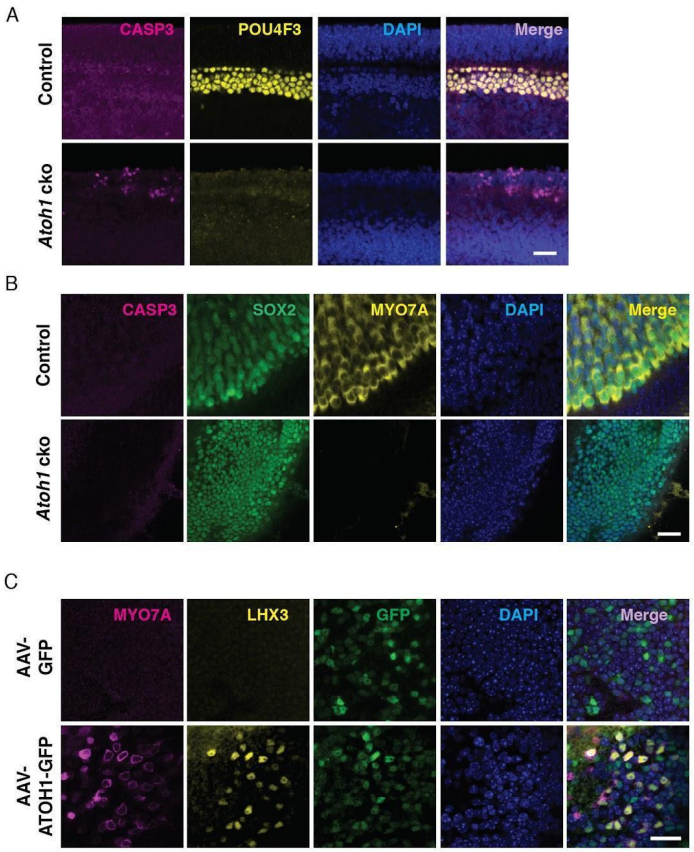

Figure S7. Exogenous ATOH1-expression is sufficient to rescue expression of hair cell genes in the *Atoh1* mutant utricle.
(A) Confocal images of the basal region of the E15.5 *Atoh1* mutant cochlea show massive apoptosis as a result of the absence of *Atoh1* expression. Scale bar, 20 μm; cKO, conditional knock out; CASP3, active caspase 3.
(B) Confocal images of the utricular sensory epithelia (P2) shows no cell death in the absence of *Atoh1* expression. The utricular progenitors remain as a monolayer of SOX2-positive cells with normal appearance. Scale bar,20 μm.
(C) Confocal images of the cultured utricle from *Atoh1* mutant mice (P2). ATOH1 overexpression (AAV) induces the expression of hair cell markers, MYO7A and LHX3, in the infected cells. Scale bar, 20 μm; AAV, Adeno-associated Virus.

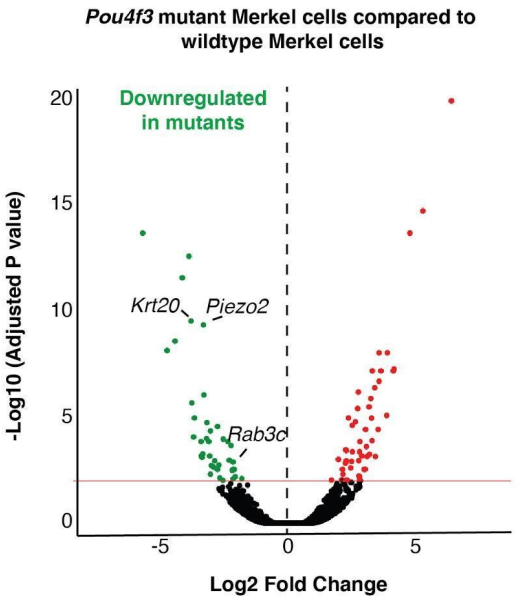

Figure S8. Gene expression comparison between wild type and *Pou4f3* mutant Merkel cells. Volcano plot shows differential gene expression between wild type and *Pou4f3* mutant Merkel cells. Marker genes for mature Merkel cells are indicated, including *Krt20*, *Piezo2* and *Rab3c*.

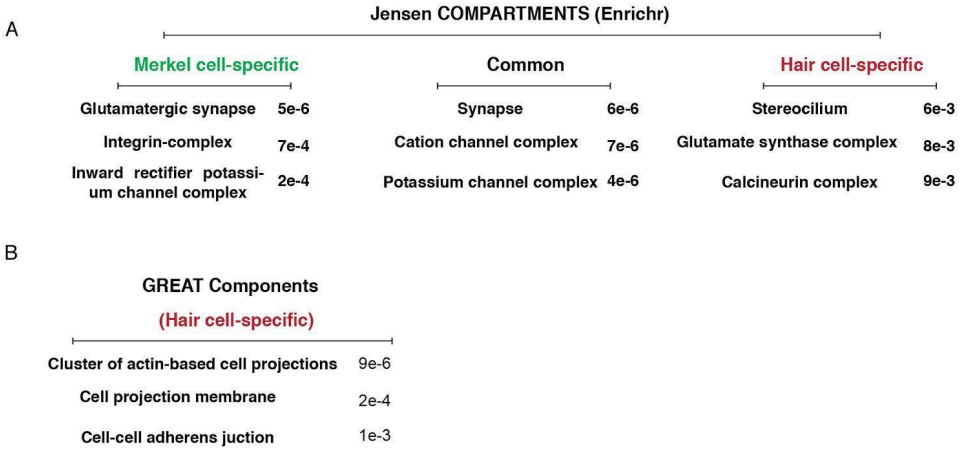

Figure S9. Gene ontology analysis of the POU4F3-dependent ATOH1 targets in hair cells and Merkel cells.
(A) GO terms from Jensen COMPARTMENTS database (87) of the POU4F3-dependent ATOH1 targets in hair cells and Merkel cells.
(B) GO term results from GREAT Components database (88) of the POU4F3-dependent ATOH1 targets in hair cells.

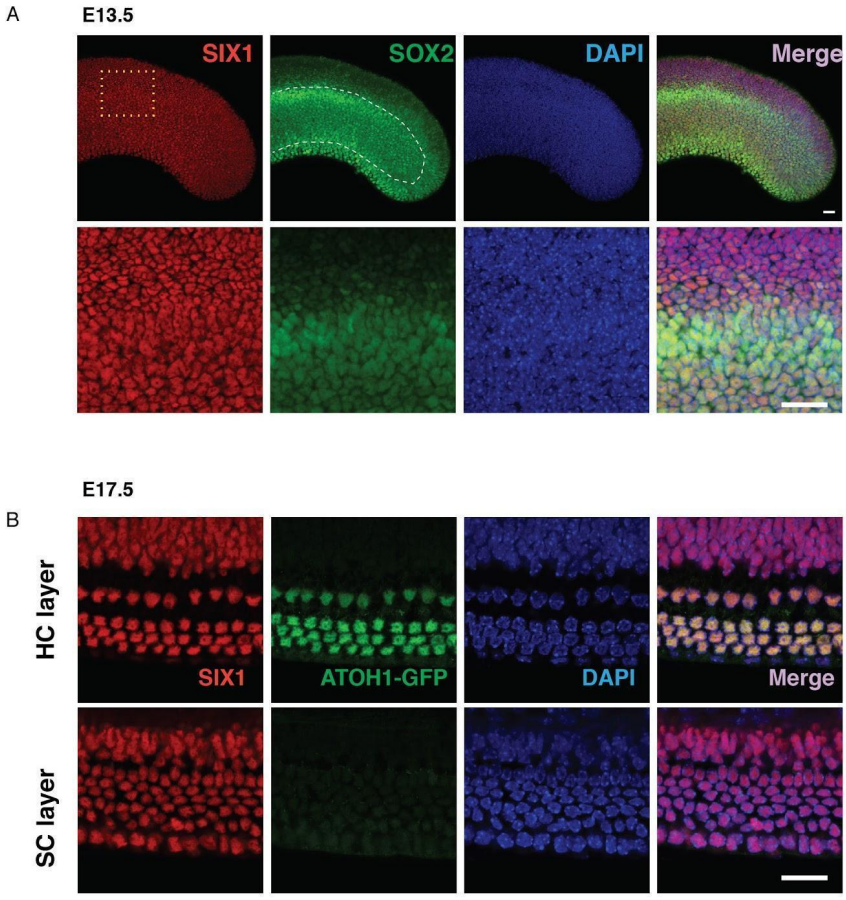

Figure S10. SIX1 is broadly expressed in the cochlea.
(A) Confocal images of low magnification (upper row) and high magnification (within dotted square) (bottom row) from the E13.5 cochlea. SIX1 is ubiquitously expressed in the cochlea. SOX2 expression marks the prosensory domain. Scale bar, 20 μm.
(B) Confocal images showing that SIX1 is expressed ubiquitously in the E17.5 cochlea. Scale bar, 20 μm.

Supplemental Datasets

Dataset S1. E17.5 hair cell specific genes. Transcriptome dataset from E17.5 hair cells was compared to E13.5 progenitor transcriptome, and genes with fold change > 4.0, FDR < 0.1, and FPKM in hair cells > 5.0 were chosen as E17.5 hair cell specific genes.

Dataset S2. Expression changes of hair cell specific genes in *Pou4f3* mutant hair cells vs wild type hair cells. Known deafness-associated genes are colored in red.

Dataset S3. Significantly down-regulated genes in *Pou4f3* mutant Merkel cells vs wild type Merkel cells. Genes with log2(fold change) < -1.5 and FDR < 0.1 are shown.

Supplementary References:

79. L. Tong, *et al.*, Selective Deletion of Cochlear Hair Cells Causes Rapid Age-Dependent Changes in Spiral Ganglion and Cochlear Nucleus Neurons. *J. Neurosci.* **35**, 7878–7891 (2015).

80. N. F. Shroyer, *et al.*, Intestine-Specific Ablation of Mouse atonal homolog 1 (Math1) Reveals a Role in Cellular Homeostasis. *Gastroenterology* **132**, 2478–2488 (2007).

81. S. Picelli, *et al.*, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

82. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

83. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

84. Y. Zhang, *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

85. F. Ramírez, *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

86. S. Heinz, *et al.*, Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

87. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

88. E. Y. Chen, *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

89. C. Y. McLean, *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

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

91. C. Trapnell, *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

92. B. Fritzsch, K. W. Beisel, Keeping Sensory Cells and Evolving Neurons to Connect Them to the Brain: Molecular Conservation and Novelties in Vertebrate Ear Development. *Brain. Behav. Evol.* **64**, 182–197 (2004).