nature portfolio

https://doi.org/10.1038/s41586-024-08577-5

Supplementary information

Repurposing of a gill gene regulatory program for outer-ear evolution

In the format provided by the authors and unedited

Supplementary Methods

Genomic coordinates for open chromatin regions tested in vivo

Human (GRCh38): chr2:205,714,322-205,715,569 (Hs_NRP2-p1); chr16:90,020,493-90,021,397 (Hs_GAS8-p1); chr7:137,316,428-137,317,390 (Hs_PTN-p1); chr15:25,504,048-25,504,926 (Hs_UBE3A-p1); chr14:91,892,906-91,893,798 (Hs_FBLN5-p1); chr18:51,614,879-51,615,771 (Hs_SMAD4-p1); chr10:13246947-13247862 (Hs_UCMA-p2); chr9:14,340,296-14,341,141 (Hs_NFIB-p1); chr11:20,051,838-20,052,829 (Hs_NAV2-p1); chr20:7,352,028-7,353,009 (Hs_BMP2-p1); chr8:118,506,525-118,507,501 (Hs_SAMD12-p1); chr4:100147709-100148800 (Hs_DDIT4L-p1); chr2:133,247,473-133,248,327 (Hs_NCKAP5-p1); chr18:8,471,234-8,471,836 (Hs_RAB12-p1); chr5:7842428-7843409 (Hs_C5orf49-p1); chr7:96833519-96834377 (Hs_DLX5/6-p1)

Zebrafish (GRCz11): chr19:41423448-41424749 (Dr_dlx5a/6a-p1)

Horseshoe crab (GCF_000517525): NW_013665731.1:904913-905717 (Lp_dll-p1)

DNA sequence for motif-mutated enhancers

Dr_gata3-p1_DLXdel:GFP

AGGGGCATTTCAAACTTATTTGTATATCATTATTATATATTATTTATGAATCAATTCCCAAAAT
AACAATATATCCAGATGTTCCTTGTCTGCTTTTATTAACTTCTTGGCCCAGTTTTATTCTAAAC
CTACATCTAAGCGCAAGCAAACACAGAACATGTTTTAGAGACCAGGCCAAGGCTGTAATTTGTT
TCAATAAATACGCTTTTAGTGATTCTCTGCTTTTCTAACCTGGTCACTGCTGCCCTCAGGCTTT
TCCTAACCAACAACTCTTTCCTGTCATTACGAGCGTGTACGAGTGTGTGCGTTATGTGTGTGAA
CTTCTTCCAGGGTTAAAAATACAAATGGTGTGCATTTTAAAAAGGCTTTTTTTTAGCCTGTGGT
GGATGCAGTGTTTTGTTTGGAAGAAAAAAATGGGATGCACGTAACTTACTTCACCGTAATAGCA
TTTAGTCCGCGCTCCTTATCTTATTTCAAAAGGCAAACTTCAGAAGGGCGGAGGTACTGTACGT
CTCCAACAAAAACACACCTTGAAGAGAGCGTGCATGCATAGAGAAACCACACACATATTAAACT
GTATGTTTACCAGATGAGTCACTAAGCATTAGAATTTGGAAGAGCTAATAGTTTAAATAGAATT
TGCAGCAAACT

"AATT” deleted at wildtype sequence position 89, 436, 456, 483, and 506. For each position where sequence has been altered from wild type the previous base pair is bolded and underlined.

Hs_NRP2-p1_DLXdel:GFP

ATTGGGACTCTTGCGCTCTGGGCCATGCATTTCTTAGTCCAGATGAGTTGTCTGATTAATGAAC
ATGGGTAGCAGAGCCTCTCAACCTTTGTGCACATTGCCAGGGCATCACCATTGCTCACAGGAAT
GTGTCTGTCCAGCAGTGCTTGCCTTAGTGTGAAAATAGCACGGAGGGGTTTTCTGGAAGGTTTG
GGAGTTTCCCCTGAGAACAAATGAAGCGTCTTTTTACTTTTGTTTTGGAAATGAGGAAGCATGT
TTTCAAAGTCAGCACCATCTCCCACACTGAAACCACAGACCTAAACAGATTCCCCAACCTCACG
GTCTTCTCAATCTGTGACCATCTCTCCATTTCTCCCTGGGGGTTGAAACTGCCATTTTGACTTC
TCCAGACTCTAAGGAGCTTCAGCGGTTCTTTTTGGACTGTGGGAGTCCAGCCCTGTGGGGTCTA
AGAATGAATTAATCCTGGCCCACTCCTTAGACTTCCTTTACAGGCCTTTGAAAGTGGCTAAGTA
ACATAGCTCCTGGGGGCTGAAGTGAGGCCTCAGGCCTCCCCAGCTCACACCAGCCACTCCTTCC
GGCAGCCGGCAACCCCAGATCCTTGTATGTGACAGGATCCTTTGGCAGGCACATCCCACAGAAG
GGTGCATTCCCAGCCCTCTCCTTACCCGCAAGGAAGAGATTTCTTAAGTGTTGCTTTAACAGGT
TGATGCAAAATAAATAAACAGAAAGGGAGGGAGAGGGAGATGGATGGTTTTCAGTAAGGTGCAG
CGGCCAGCAGTTGGAAGGGAGCTGGTGGGAGGAAGTTTCCCACCTAAGCAGTTGTGACAGGAGT
ATTAACTGGGAGCCCAAAAGCAGCACAAACGCAGCACAGTGAGGACTGCCAGAGCACATACCGG
GGCTGGAAAATGTGTGTCTGTCAAGATAAAATGGTGCGTTTGCACCTTCCCATGAACGGAGCCA
AGGGTATCTTTGAGGTGCCTGCTTCCCAGCCCTGGCCAGTTTGACTCTGCTTCAAATGAGCACA
GGCCTCCACTTCAGGAGGGCAGCCTGCCCTGCCTCTCACTGAGGCAGTTGTGTGCACGTAGACA
GTGGGCCCCTGGAGACTCACCCCTGCTTTTGGAGAAAGGTTTTTATCGCCTTATCTAACATGAG
GCTGATTCCTTTTAAGAACACATGTTGACTAGACACTGCTGATGGGAACCTCGCAGAGTCATCT
TAGGGAGGCGGTG

“ATTA” deleted at wild-type sequence positions 496 and 944, “ATTAT” deleted at position 767, and “TAAT” deleted at position 1085. For each position where sequence has been altered from wild type the previous base pair is bolded and underlined.

Additional protocol details for mouse transgenesis

The *Dr_ucmaa-p1* and *Dr_dlx5a/6a-p1* enhancers were PCR-amplified from zebrafish genomic DNA and cloned into the PCR4-Shh::lacZ-H11 plasmid (Addgene #139098) upstream of the minimal Shh promoter and lacZ reporter gene. Pronuclear microinjection of B6D2F1 zygotes was performed using 25 ug of plasmid, 50 ng/uL of the published gRNA targeting the H11 locus (iDT Alt CRISPR system) and 20 ng/uL Cas9 protein (IDT cat. no. 1074181). Injected zygotes surviving to the 2-cell stage were implanted into pseudo-pregnant females. Positive integration of the zebrafish enhancer was confirmed by PCR genotyping using primers designed to amplify the insert followed by Sanger sequencing of the PCR product.

Additional protocol details for human tissue dissociation

Using scissors and a scalpel, the tissue was finely minced and dissociated by mechanical and enzymatic digestion: the tissue was incubated for 4 h with gentle agitation at 37^{o}C in conical flasks containing 1 mg/mL collagenase II (Worthington), 1 mg/mL dispase (Gibco), 1:1000 gentamicin (Teknova), 1:500 primocin (Invivogen), 1% penicillin/streptomycin/amphotericin B (P/S/A; Corning), and 10% fetal bovine serum (FBS; Corning) in DMEM/F12 (Corning). The mixture was pipetted in the sterile hood every 30 min. The resulting cell suspension was passed through a 70-micron filter (Fisher) and washed in 2% FBS/PBS solution to stop enzymatic digestion. For red blood cell lysis, the cells were incubated in a solution consisting of 1 mL 2% FBS/PBS and 9 mL ammonium chloride solution (STEMCELL Technologies; cat. no. 07800) for 10 minutes on ice and washed in 2% FBS/PBS. We then resuspended the remaining cells in 10% dimethyl sulfoxide (DMSO) and 20% FBS in DMEM/F12 medium and cryopreserved at -

80ºC for one month. For human fetal outer ear replicate 1 and nose tissues, we additionally performed negative selection using the magnetic activated cell sorting system (Miltenyi Biotec) using a primary anti-CD45 monoclonal antibody, PE (eBioscience, 12-9459-42), and an anti-PE secondary antibody (Miltenyi, 130-048-801) to remove CD45+ white blood cells as per manufacturer's instructions, just prior to cryopreservation.

Additional protocol details for human multiomics data analysis

Cell Ranger ARC-output count matrices in H5 format of individual libraries were used to create Seurat objects by "CreateSeuratObject" and "CreateChromatinAssay" functions (EnsDb.Hsapiens.v86 used to create Granges containing annotations for the human genome). For analysis of human nose and outer ear replicate 1, separately generated Seurat objects were merged using the "merge" function prior to downstream processing. For simultaneous analysis of all 4 human tissue libraries (Extended Data Figs. 5 and 6), Cellranger aggr function was used to merge all four libraries and normalize read depth across all cells prior to generating a Seurat object. In both cases, quality control was performed by setting thresholds of accessible region counts "nCount_ATAC", transcript counts "nCount_RNA" and ratio of mononucleosome to nucleosome-free region "nucleosome_signal" to remove outliers that may represent low quality cells or doublets. RNA data were normalized and the top 2000 most variable genes selected using "FindVariableFeatures" for scaling. Dimensional reduction was performed using "RunPCA" function with the number of principal components selected by ranking of components based on percent of variance explained by each. Data were then subjected to neighbor finding ("FindNeighbors") and clustering ("FindClusters", resolution=0.6 for

outer ear 1 and nose merged, resolution=1.0 for all four libraries merged). The data were visualized through UMAP. Cell type identification was performed based on differentially expressed genes across clusters generated from snRNAseq data (full gene lists provided in Table S1). Peak calling was performed using “CallPeaks” Signac function, which uses the MACS2[61] algorithm. Peaks on nonstandard chromosomes and in genomic blacklist regions were removed using “keepStandardChromosomes” (pruning.mode=coarse) and “subsetByOverlaps” (using blacklist_hg38_unified). Counts in each peak remaining in the peak set were quantified for the merged multiome datasets using FeatureMatrix function and the resulting counts were added as a “Peaks” ChromatinAssay to the Seurat object. We used “FindAllMarkers” to detect genes with highly enriched expression and ATAC peaks with highly differential accessibility in each cluster generated from the above-described dimensionality reduction analysis of snRNAseq data.

For determining the number of ATAC peaks differentially accessible between outer ear and nose cartilage and between epiglottis and nose cartilage, we used the “FindMarkers” function in the Seurat R package with a logfc.threshold of 1 and only.pos=T, and filtered to include features with pct.1 > 0.1 (accessible in greater than 10% cells from identity 1). We then found the intersection of these to determine accessible regions differentially accessible in both types of elastic cartilage relative to nose hyaline cartilage (“intersect” in dplyr R package). To identify outer ear- and epiglottis-enriched features that are differential with hyaline cartilage, we used the “setdiff” function in the dplyr package.

Additional details on selection and analysis of human elastic cartilage DARs

Outer ear-specific DARs were selected from a list of regions generated by comparing outer ear and nose cartilage clusters from the analysis of merged outer ear replicate 1 and nose multiome datasets. Out of 10 outer ear DARs tested, 4 were selected for proximity to genes with roles in cartilage and/or elastin biology (*Hs_FBLN5-p1*, *Hs_UCMA-p1*, *Hs_UCMA-p2*, *Hs_SMAD4-p1*), 2 were selected for proximity to genes differentially expressed in elastic cartilage (*Hs_NRP2*, *Hs_PTN-p1*), 1 was selected for association with a transcription factor whose motif was the top-most enriched in elastic cartilage (Hs_NFIB-p1; **Extended Data Fig. S2i**) while the other 3 were randomly selected from the top DARs with no bias (*Hs_UBE3A-p1*, *Hs_NAV2-p1*, *Hs_GAS8-p1*). Out of 6 nose DARs tested, 3 were selected for known roles in cartilage biology and/or differential expression in nose cartilage (*Hs_BMP2-p1*, *Hs_DDIT4L-p1*, *Hs_NCKAP5-p1*), and the other 3 were chosen from the top DARs without bias (*Hs_RAB12-p1*, *Hs_SAMD12-p1*, *Hs_C5orf49-p1*). For comparison with epiglottis, we subset the Seurat clusters "cartilage 1", "cartilage 2", "cartilage 3", and "cartilage 4" from the merged analysis of all four human fetal tissue multiome datasets and made coverage plots of for genomic regions containing each of the analyzed DARs.

Additional protocol details for horseshoe crab nuclei dissociation

From the adult male horseshoe crab, the opisthoma was separated and book gills dissected out using shears. The exoskeleton was peeled off to reveal the cartilage tissue. The tissue was cut using scissors into 1 cm X 1 cm pieces. Each piece was flash frozen in isopentane cooled in a liquid nitrogen bath and immediately transferred into a cryovial. The flash frozen tissue was stored at -80^{o}C for two weeks. Subsequently, 200 mg of tissue

was weighed out and finely minced using ice-cold scissors. The minced tissue was transferred into ice-cold homogenization buffer (250 mM sucrose, 10 mM Tris pH 8.0, 25 mM KCl, 5 mM $MgCl_2$, 0.1% Triton X-100, 0.5% RNAse inhibitor (Sigma; cat. no. 3335399001), 1X protease inhibitor (Roche; cat. no. 5056489001), and 0.1 mM dithiothreitol (DTT; Sigma cat. no. 646563)). Nuclei were released through mechanical digestion: a chilled dounce was used to release nuclei by 10 strokes of the loose pestle and 7 strokes of the tight pestle from the minced tissue. The dounced nuclei were strained, filtered, and washed by centrifugation at 1000 g for 10 min before fluorescence-activated nuclei sorting.

Additional protocol details for horseshoe crab multiomics data analysis

Cell Ranger ARC-output count matrices in H5 format of individual libraries were used to create Seurat objects by “CreateSeuratObject” and “CreateChromatinAssay” functions. For the latter, GCF_000517525.1_Limulus_polyphemus-2.1.2_genomic.gtf from NCBI RefSeq was used to create a custom annotation for the horseshoe crab genome. Quality control was performed by setting the thresholds of accessible region counts (nCount_ATAC) < 15000, transcript counts (nCount_RNA) < 2500 and ratio of mononucleosome to nucleosome-free region (nucleosome_signal) < 1.5. Factor level names needed to be altered in the Seurat object to remove the “_” in genome contig names. Owing to high levels of ambient RNA and ATAC fragments in the dataset, we used DecontX[62] and celda (CELLular Latent Dirichilet Allocation; 1.16.1)[63] packages on snRNAseq and snATACseq counts to estimate and remove contamination in individual cells and added the decontaminated counts back to our Seurat object as “decontXRNA”

assay and “decontXATAC” ChromatinAssay respectively. decontXRNA data were normalized and the top 2000 most variable genes selected using “FindVariableFeatures” for scaling. Dimensional reduction to 50 principal components was performed using “RunPCA”. Data were then subjected to neighbor finding (“FindNeighbors”, k=30) and clustering (FindClusters, resolution=0.6). Peak calling was performed on the decontXATAC ChromatinAssay using “CallPeaks”, which uses the MACS2 algorithm. Counts in each peak remaining in the peak set were quantified for the multiome dataset using “FeatureMatrix” and the resulting counts were added as a “Peaks” ChromatinAssay to the Seurat object. We used “FindAllMarkers” to detect genes with highly enriched expression and ATAC peaks with highly differential accessibility in each cluster generated from the above-described dimensionality reduction analysis of snRNAseq data. The differentially expressed genes were used to assign cell type identity based on known conserved markers for specific cell types.

Additional protocol details for whole mount X-gal staining

Briefly, pregnant mothers from timed matings were euthanized and embryos dissected from their uterus. The embryos were rinsed in ice cold PBS with 2 mM $MgCl_2$ to prevent loss of ß-galactosidase activity. Following a fixation in fixative solution (0.05% glutaraldehyde, 1% paraformaldehyde (PFA), 100 mM phosphate buffer, 5 mM EGTA, 0.1% NP-40 and 2 mM $MgCl_2$) for 2 h at 4^{o}C, the embryos were immersed in X-gal staining solution (0.1 M phosphate buffer, 0.01% sodium deoxycholate, 0.02% NP-40, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 20 mM Tris-HCl, 2 mM $MgCl_2$ and 1 mg/mL X-gal (Thermo; cat. no. B1690)) for 24-48 h at 25^{o}C. The embryos were rinsed

in PBS, fixed in 4% PFA, and cleared in 1:1 solution of CUBIC-1 (25% *N*,*N*,*N*′,*N*′-Tetrakis(2-hydroxypropyl)ethylenediamine, 25% urea, 15% Triton X-100 in distilled water) and PBS followed by 100% CUBIC-1 for 1 day to 1 week at 37°C.

Additional protocol details for Miller's staining

Miller's stain was prepared as follows: 200 mL of boiling distilled water was added to 1 g each of Victoria Blue 4R (City Chemical LLC, cat. no. 554D), New Fuchsin (VWR; cat. no. IC15582325) and Crystal Violet (Sigma-Aldrich cat. no. C0775) and mixed for 1 min with a spatula. To this were added 4 g each of resorcin and dextrin. Next, 50 mL of freshly prepared 30% ferric chloride solution was added. The mixture was boiled for 5 min on a hot plate and subsequently filtered through Whatman's filter paper before the volume was made up to 200 mL with 95% ethanol. 4 mL concentrated hydrochloric acid was added to the mixture. Miller's elastic staining was performed on paraffin-embedded sections as follows: slides were baked at 60°C for 1 h followed by deparaffinization and re-hydration to distilled water. Slides were then stained in potassium permanganate (1% (w/v) aqueous solution) for 1 minute, rinsed in distilled water and decolorized in 1% (w/v) oxalic acid until colorless (between 30 s and 1 min). The slides were rinsed under running tap water, then immersed in 95% ethanol for 1 min followed by staining with Miller's stain in a closed vessel for 3 h. The slides were rinsed in 95% ethanol by briefly dipping, then washed in running water and moved into Curtis' Picro-Ponceau staining mixture (1% (w/v) aqueous Ponceau S stain and 1% (v/v) glacial acetic acid in saturated aqueous picric acid) for 4 min. Finally, the slides were rinsed twice in 95% ethanol, once in 100% ethanol, then cleared in xylene and mounted for imaging.

Additional protocol details for Alcian Blue staining on sections

Alcian Blue solution was prepared at using Alcian Blue, 8X (Sigma; cat. no. A5268) in 3% acetic acid solution adjusted to a pH of 2.5 with additional acetic acid. 0.1% Nuclear Fast Red (Sigma; cat. no. N8002) was prepared in 5% aluminum sulfate (Sigma; cat. no. A7523) by adding nuclear fast red slowly to the aluminum sulfate solution, slowly heating to boil, followed by cooling and filtering. A grain of thymol (Sigma; cat. no. T0501) was added as a preservative. Alcian blue staining was performed on paraffin sections from various fishes as follows: slides were baked and then deparaffinized by clearing in xylene for 3 min twice, followed by hydration to distilled water. Slides were incubated in 3% acetic acid for 3 min, then stained in Alcian Blue solution for 30-45 min. The slides were rinsed in running tap water for 2 min and then counterstained in nuclear fast red solution for 3-5 min. Following a second rinse in running tap water for 1 min, the slides were washed once in distilled water (2 min), dehydrated in graded ethanol solutions, cleared in xylene, and mounted for imaging.

Additional protocol details for imaging

Fluorescent images of section in situ hybridizations and live transgenic fish were captured on a Zeiss LSM800 confocal microscope using ZEN software. Additionally, to acquire highly magnified color images of dissected gills stained with Alcian blue, and Van Gieson Verhoeff-stained zebrafish sections, the Zeiss LSM800 confocal microscope was used for sequential scanning with red, green, and blue lasers. A Zeiss AxioScan.z1 slide scanner with a 20X 0.8NA objective was used to image paraffin sections of whole mount X-gal

stained *Dr_dlx5a/6a-p1:lacZ* E15.5 mouse embryos and adult green anoles. For imaging all other X-gal-stained mouse sections, Miller's stained paraffin sections of adult zebrafish and mouse outer ear, and Trichrome stained horseshoe crab sections, a Leica DM2500 compound microscope was used. X-gal-stained whole-mount mouse embryos were imaged in brightfield using a Leica S8APO stereo microscope. We took advantage of the fluorescent properties of the reaction product to image whole mount X-gal stained E15.5 dpc mouse embryo sections expressing *Dr_dlx5a/6a-p1:lacZ*. Imaging was performed on a Leica DMi8 SP8 equipped with White Light Laser (WLL) and Hybrid Detectors (HyD). Images were acquired with a 25X, 0.95NA water objective the confocal pinhole set to 3.0 Airy Units. The excitation wavelength was set to 580 nm with a collection window between 603-629 nm. To eliminate autofluorescence, we temporally filtered emitted light between 5.84 and 9.34 nanoseconds. Briefly, the WLL produced pulsed laser light which allowed temporally discrete collection of emitted fluorescence. Since autofluorescence typically has a short fluorescence lifetime (the period between photon absorption and fluorescence emission), we temporally gated the earliest arriving photons following an excitation pulse. Transient transgenic Xenopus NF42 and NF47 tadpoles were imaged on a Zeiss AxioZoom.v16 wide-field fluorescent microscope. NF47-49 tadpoles and whole-mount immunostained fixed tadpoles were imaged on a Zeiss LSM800 confocal microscope.
The adult horseshoe crab was photographed using a Google Pixel 4a phone camera and the green anole head was photographed using a Nikon D60 camera.

Supplementary References

61 Gaspar, J. M. Improved peak-calling with MACS2. *BioRxiv*, 496521 (2018).

62 Yang, S. *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome biology* **21**, 1-15 (2020).

63 Wang, Z. *et al.* Celda: A Bayesian model to perform co-clustering of genes into modules and cells into subpopulations using single-cell RNA-seq data. *NAR Genomics and Bioinformatics* **4**, lqac066 (2022).