nature portfolio

Corresponding author(s): J. Gage Crump

Last updated by author(s): Feb 14, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	Cellranger Arc v2.0.0 (10X Genomics) was used for alignment and peak calling to generate a peak-by-cell and gene-by-cell count matrix. Multiomics data were processed by Seurat (4.3.0.1) and Signac (1.10.0)48 packages in RStudio (R version 4.3.1). We used the python package SAMap (version 1.0.15) to correlate gene expression patterns and determine cell type homology. In RStudio, we used the taRifx (1.0.6.2)50 and SeuratDisk (0.0.0.902) package to convert raw count matrices of zebrafish and human scRNAseq/snRNAseq Seurat objects to h5ad format and to transfer cell type annotations to this format. Next steps were performed in Jupyter Notebook (version 6.5.4) using Python (version 3.9.18). We used "CallPeaks" from Signac to extract accessible peak coordinates from both elastic and hyaline cartilage clusters. HOMER v4.9.1 was used for de novo motif discovery. Other packages and software used for analysis include sci-kit-learn library, bedtools2 v2.3.0, blastn v2.11.0, MACS2, dplyr v1.1.4, celda v1.6.1, DecontX v1.4.0. Code for the sequence conservation analysis pipeline and random forest classifier are available at Github [https://github.com/MathiThiruUSC/Ancient_gill_origin_outer_ear] as provided in Code Availability section of the manuscript. Zeiss ZEN2.3 BLUE software was used for confocal image acquisition and processing.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Single-nuclei multiomics data for human fetal outer ear, nose, and epiglottis tissues can be accessed at GEO GSE255677, and single-nuclei multiomics data for adult Limulus polyphemus book gill tissue at GEO GSE255678. The following reference genomes utilized in this study are publicly available: human GRCh38 (https://www.ensembl.org/Homo_sapiens/Info/Index), zebrafish GRCz11 (http://www.ensembl.org/Danio_rerio/Info/Index), green anole AnoCar2.0v2.11 (https://www.ensembl.org/Anolis_carolinensis/Info/Index), Xenopus tropicalis UCB_Xtro_10.0 (https://www.ensembl.org/Xenopus_tropicalis/Info/Index), and horseshoe crab GCF_000517525 (https://www.ncbi.nlm.nih.gov/assembly/91501).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	No data on population characteristics was collected/used.
Recruitment	A 22-week gestation fetus was obtained from the University of California Los Angeles Center for AIDS Research Gene and Cellular Therapy Core using institutional review board-approved de-identified and consented electively aborted human fetuses.
Ethics oversight	USC and UCLA IRB.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample sizes were calculated. We generally aimed to identify n=3 individual recorded occurrences for each experimental result reported. In all cases, the research is qualitative rather than quantitative. In each case we are reporting the presence or absence of signal (transgenic activity, mRNA expression, or color development in a histological stain) within a group expected to be relatively homogeneous for the trait being investigated (gene expression or tissue phenotypic features in wild-type animals or reporter activity in a group of transgenic animals derived from a founder individual that experienced a single genomic integration event).
Data exclusions	No data was excluded
Replication	For all transgenic constructs, we identified and characterized at least 2 independent founder alleles and successfully noted reproducible expression patterns as reported in the manuscript in at least 3 independent animals per line. The few exceptions have also been stated in the text: for Hs_RAB12-p1:GFP in zebrafish, we noted two distinct expression patterns as stated in the text. For Hs_SAMD12-p1:GFP in zebrafish, we were unable to generate a second founder allele. For Dr_dlx5a/6a-p1:lacZ in mouse, we identified only one founder with weak expression in the ear region and strong neural tube expression whereas the second founder only weakly recapitulated the neural tube expression. We associated this result with the weak activity of this enhancer, which was consistent with its weak activity when tested in zebrafish. For all in situ hybridization and other stains performed on sections, the stain was repeated on at least 2 individual animals. Exceptions were made in the case of rare or difficult to generate samples such as juvenile horseshoe crabs. In these cases, the result was either re-confirmed on multiple comparable sections on different slides and/or technical replicates were performed on adjacent sections at least 2 additional times to ensure as high a level of reproducibility as possible.
Randomization	To ensure that the technical covariates of sequencing depth and batch effect were minimized in our single cell analyses, we made sure that

Randomization	datasets to be subjected to direct comparison with one another were generated at the same time, pooled together following separate barcoding, and sequenced in a single run. When datasets that were separately sequenced were to be compared, we utilized Cellrangers aggr function, which normalizes for effective sequencing depth by subsampling reads.
Blinding	The samples profiled for single-cell analysis and transgenic lines imaged were known and not blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	All zebrafish were generated on a Tubingen background. Experimental zebrafish included those aged 1.5 dpf, 3 dpf, 4 dpf, 5 dpf, 6 dpf, 14 dpf, 17 dpf, 2 mpf, and 3 mpf. Mice were maintained on a C57BL6/J background and experimental mice were utilized at embryonic stages E13.5, E15.5, postnatal day P0, and 6 weeks post-natal. Xenopus laevis tadpoles aged NF42, NF47, and NF49 were analyzed. Green anoles were harvested at adult stages. Mice were raised in 14-hour light/ 10-hour dark cycle at approx. 50% humidity and 21C.
Wild animals	An Atlantic horseshoe crab was collected from the wild by Marine Biological Laboratories and shipped overnight at the approximate age of 8 years post-fertilization in a container filled with moist grass. A separate juvenile horseshoe crab approximately 1 year post-fertilization also wild-caught at the Marine Biological Laboratories was transported in a container filled with sand and seawater. Upon arrival, the animals were euthanized by submerging in filtered seawater prepared at a salinity of 1.025 using Instant Ocean salt containing 1g/L Tricaine-S (Syndel) and buffered with 4.5 mM sodium carbonate at 25oC until the gills stopped moving. They were then utilized in a single nuclei multiome study and for in situ hybridization analysis respectively.
Reporting on sex	Experiments conducted on zebrafish before gender is determined. For mouse embryo research, gender was undetermined.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	The Institutional Animal Care and Use Committee of the University of Southern California approved all animal experiments (Protocol 20771).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A