Spherical harmonics analysis reveals cell shape-fate relationships in zebrafish lateral line neuromasts

Madeleine N. Hewitt[1,2], Iván A. Cruz[3], David W. Raible[1,2,3,*]

[1]Molecular and Cellular Biology Graduate Program, University of Washington School of Medicine, Seattle WA 98195, USA
[2]Department of Otolaryngology-HNS, University of Washington School of Medicine, Seattle WA 98195, USA
[3]Department of Biological Structure, University of Washington School of Medicine, Seattle WA 98195, USA

*Corresponding author: David W.
Raible draible@uw.edu

Summary Statement

Spherical harmonics analysis of three-dimensional cell shapes allows quantitative comparison of cell types and cell fate.

Abstract

Cell shape is a powerful readout of cell state, fate, and function. We describe a custom workflow to perform semi-automated, 3D cell and nucleus segmentation, and spherical harmonics and principal components analysis to distill cell and nuclear shape variation into discrete biologically meaningful parameters. We apply these methods to analyze shape in neuromast cells of the zebrafish lateral line system, finding shapes vary with cell location and identity. The distinction between hair cells and support cells accounted for much of the variation, which allowed us to train classifiers to predict cell identity from shape features. Using transgenic markers for support cell subpopulations, we found that subtypes had different shapes from each other. To investigate how loss of a neuromast cell type altered cell shape distributions, we examined *atoh1a* mutants that lack hair cells. We found that mutant neuromasts lacked the cell shape phenotype associated with hair cells, but did not exhibit a mutant-specific cell shape. Our results demonstrate the utility of using 3D cell shape features to characterize, compare, and classify cells in a living, developing organism.

© 2024. Published by The Company of Biologists Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed

Introduction

Cell shape emerges from a complex interplay between fate, function, and external forces (Chan et al., 2017). Quantifying cell shape during development has provided mechanistic insight into morphogenetic processes (Perez-Vale and Peifer, 2020) cell division and death (Chen et al., 1997) as well as differentiation (Watt et al., 1988; Kilian et al 2010). However, most studies of cell shape *in vivo* have used descriptive features (such as volume and sphericity), which may not capture subtle or multifactorial variations in shape between cell types, especially in tissues with complex, 3D organization. More complete 2D shape representations have been used to study cell motility and migration (Pinkus and Theriot, 2007; Keren et al., 2008) but these methods cannot be easily generalized to 3D shapes. To fully appreciate the diversity of cell shape states during development and understand the relationship between cell shape and fate, robust 3D cell shape analysis methods must be applied to living, developing organisms.

Recently, several 3D cell shape representations have shown promising results. One approach uses spherical harmonics (SH), where the surface of each cell is mapped to a sphere, and the SH transform is used to generate a list of coefficients that describe the cell's shape (Ruan and Murphy, 2019). Principal components analysis (PCA) can then be applied to reduce dimensionality and facilitate interpretation of cell shape variation within and between cell populations. SH and PCA has been used to understand variation in cell and nuclear shape of IPS cells in culture (Viana et al., 2023), describe organ shape trajectories in murine limb and heart development (Dalmasso et al., 2022) and 3D shapes in medical imaging (Gerig et al., 2001). These methods maintains enough information to accurately reconstruct 3D cell shapes, enabling sophisticated, data-driven analysis. Here we test the idea that we can use SH coefficients to cluster and classify cell types.

The zebrafish lateral line system is an ideal model for studying the relationship between cell shape, location, and fate in a living, developing animal. Organs of the lateral line, known as neuromasts, are small, tightly packed epithelial rosettes with apical-basal polarity, radially organized cell types, and mirror symmetry (Thomas et al., 2015). Mechanosensory hair cells (HCs) are located in the center of the neuromast and surrounded by nonsensory support cells (SCs). SCs can be distinguished based on their locations, gene expression patterns, and propensity for HC replacement during regeneration, and identified in live animals using established knock-in lines (Lush et al., 2019; Thomas and Raible, 2019). The superficial location of neuromasts and optical transparency of zebrafish larvae facilitate high resolution live imaging.

In this study, we developed a workflow to segment neuromast cells and nuclei and analyze their 3D shapes using SH and PCA. We model cell and nucleus shape using just 8 cell shape and 4 nucleus shape principal components (PCs). We found that HCs were clearly distinct from SCs in shape space, while SCs appeared to vary continuously in shape. Unsupervised clustering revealed that markers for SC subpopulations were enriched within certain clusters, suggesting these cell types have distinct cell shape features. Analysis of mutants in *atoh1a*, encoding a conserved transcription factor necessary for HC differentiation (Bermingham et al., 1999; Millimaki et al., 2007) revealed the loss of the HC shape cluster and expansion of other cell shape clusters. We also successfully built classifiers to predict HC identity from cell and nucleus shape features. Our work demonstrates that cell shape parameters can be used to characterize and classify cell types in neuromasts, laying the foundation for future studies of form, fate, and function in sensory epithelia.

Results

A Semi-automated Workflow Accurately Segments Zebrafish Lateral Line Cells and Nuclei

Neuromasts are epithelial organs with tightly packed, apically constricted cells that present challenges for imaging and segmentation (Figure 1A). For all experiments, we used the zebrafish *Tg(cldnb:LY-EGFP)* transgenic line that labels neuromast cell membranes (Haas and Gilmour, 2006). In addition, some zebrafish contained knock-in alleles that labeled cell populations of interest, for analysis described below. For numbers of neuromasts and cells from each line, refer to Supplemental Table 1. All fish were stained with the far-red nuclear dye DRAQ5 (Smith et al., 1999) Fish were imaged live at 5 days post fertilization (dpf), when neuromasts are functionally mature.

We trained machine learning models to detect cell boundaries from membrane GFP signal and nuclei masks from DRAQ5 signal, and used model results as the foundation for instance segmentation. We first applied a distance transform watershed algorithm to segment every nucleus. Labeled nuclei were then used as seeds for marker-based watershed of membrane boundary predictions. The accuracy of segmentation was further improved through proofreading and manual correction. This semi-automated workflow produced good quality segmentations for most cells (Figure 1A). In total, 3274 cells from 49 different neuromasts across 6 different imaging sessions were segmented.

Spherical Harmonics Accurately Represent Most Neuromast Cell Shapes

To quantitatively compare cell shapes, we used a method based on SH expansion (Viana et al., 2023). The result is a vector of coefficients that describe cell shape (organized into a cell-by-coefficients table), on which the inverse transform can be applied to create a reconstruction of the cell (Figure 1D). To minimize variation due to cell orientation so that we could query intrinsic cell shape, we corrected for (1) the radial symmetry of cells around the apical-basal axis of the neuromast, (2) differential tilt along the radial axis as cells converge at the apical end of the rosette, and (3) tilt introduced in image acquisition resulting in the z-axis slightly off parallel to the apical-basal axis. An example of a cell aligned by this procedure is shown in Figure 1B,C.

Following alignment, we applied SH expansion to generate 2178 SH coefficients for each cell (Figure 1D). To measure reconstruction fidelity, we used directed Hausdorff distance. We found a bimodal distribution of error, with 84% of cells accurately reconstructed (Supplemental Figure 1A). Inaccurately reconstructed cells tended to be located midway between the center and periphery of the neuromast (Supplemental Figure 1B) and had shapes that rendered them unsuitable for SH expansion, such as high concavity (Supplemental Figure 1C). These cells tended to have greater cell surface area and height (Supplemental Figure 1D, top). However, nuclei of excluded cells had similar dimensions to those of retained cells (Supplemental Figure 1D, bottom). In sum, some tall, curved cells located in a ring around the middle of the neuromast were not reconstructed well by SH (421 cells, representing 16% of the original dataset), and were excluded from subsequent cell shape analysis.

Principal Components Analysis Reveals Variations in Cell Shape Between Neuromast Cell Types

To identify the main axes of shape variation, we applied principal component (PC) analysis. The 2178 SH coefficients were reduced to 8 PCs representing approximately 70% of the total variance (Supplemental Figure 2). A key advantage of our approach is that from any point in this 8-dimensional shape space, it is possible to calculate the corresponding SH coefficients and create 3D representations of real or theoretical cells. We wondered whether variation in shape between cell types would be continuous (indicating smooth transitions between different shapes) or discrete (suggesting some populations have features that set them more clearly apart from other cells). To visualize shape variation, each PC was first individually z-scored to generate “shape modes.” We used the point corresponding to the origin in shape space (where the value of all 8 cell shape modes is 0) to create a visual representation of the mean cell shape. We then varied the value of one shape mode while keeping other shape modes constant (Figure 2A). To determine how each shape mode varied with position across neuromasts, we classified cell position by corresponding shape mode value (Figure 2B). A small number of interpretable parameters therefore captures the majority of cell shape variation.

To test whether HCs would be quantitatively distinct from SCs, we manually annotated HCs in our dataset using the presence of apical stereocilia, which are visible in the *Tg(cldnb:LY-EGFP)* transgenic line. We found that cell shape mode 1 (CSM1, representing 27% of total variance) appeared to represent “HC-ness.” Positive CSM1 scores corresponded with greater cell volume and width (Supplemental Figure 3B), as well as a more convex, flask-like shape (Figure 2C). Cells with positive CSM1 scores also tended to be closer to the neuromast center (Figure 2D, Supplemental Figure 3D), consistent with the characteristic HC location. This result suggests that much of cell shape variation in neuromasts is due to differences between HCs and SCs.

In contrast to CSM1, cell shape mode 2 (CSM2; 16% of total variance) did not distinguish between HCs and SCs; rather it appeared to distinguish among SC types. Cells with increasing CSM2 scores were progressively taller and less deep (Figure 2E, Supplemental Figure 3B). CSM2 was also negatively associated with being located toward the neuromast periphery, particularly for SCs (Figure 2F, Supplemental Figure 3D). HCs tended to have average CSM2 scores, clustered around zero, and showed a narrower distribution than SCs (Supplemental Figure 3C). Taken together, the relationships between cell type and cell position likely explain the “bullseye” pattern observed when CSM2 scores are plotted by relative cell location: SCs have higher CSM2 scores as they become more central, while HCs (which are also centrally located) have CSM2 scores close to zero (Figure 2F).These observations suggest that CSM2 represents shape variation between SCs in a way that is distinct from the variation between SCs and HCs.

To further query the relationships between shape modes and cell type, we compared shape modes to geometric features and location parameters (Supplemental Figure 3). We correlated CSM1 and CSM2 to cell width, depth, height, volume, surface area and cell distance from the neuromast center (Supplemental Figure 3B,D). We found a positive correlation between CSM1 and cell volume and cell width, and a negative correlation with distance from the center of the neuromast. For CSM2, there was a positive correlation with cell height and a negative correlation with cell depth and distance from center. HCs consistently segregated with CSM1 but not CSM2 (Supplemental Figure 3C,D). Beyond CSM1 and CSM2, other shape modes appeared to represent more subtle ways in which neuromast cell shape varied (Supplemental Figure 2).

Nuclear Shape Variation Is Captured by SHE and PCA

We also performed SHE and PCA on nucleus shape. Nuclei were aligned to account for position within the neuromast using a series of rotations similar to (but independent of) those used to align cells. Because nuclei had a much lower reconstruction error compared to cells (Supplemental Figure 1A), we did not to exclude any nuclei from analysis. Nuclear SHE coefficients were reduced to 4 principal components (PCs) representing approximately 76% of the total variance in nuclear shape

(Supplemental Figure 4). As for cell shape analysis, nuclear shape PCs were z-scored to describe them as nuclear shape modes.

Nucleus shape mode 1 (NSM1, representing 40% of total variance, Figure 3A) and nucleus shape mode 2 (NSM2, representing 18% of total variance, Figure 3C) captured distinct aspects of variation in nuclear shape. Increasing NSM1 scores were associated with a shift from short, wide shapes to a tall, crescent-like morphology (Figure 3A). NSM2 appeared to describe a progression from tall and cylindrical to flattened, rectangular shapes (Figure 3C). Nucleus shape modes 3 and 4 are summarized in Supplemental Figure 4. Comparing to other nuclear shape parameters (Supplemental Figure 5), we found NSM1 was negatively correlated with nucleus width and positively correlated with nucleus height (Supplemental Figure 5B,D). NSM2 was positively correlated with nucleus depth and negatively correlated with nucleus volume (Supplemental Figure 5B,D).

We asked to what degree nuclear shape modes distinguished among HCs and SCs. We found that NSM1 had a bimodal distribution (Supplemental Figure 5C) and was correlated with distance from the neuromast center (Figure 3B, Supplemental Figure 5D). Like CSM1, NSM1 also appeared to distinguish between HCs and SCs, but to a lesser degree (Supplemental Figure 5C). NSM2 showed a slightly skewed distribution and a correlation with distance to the neuromast center (Figure 3D, Supplemental Figure 5D), with some distinction between HCs and SCs. Nuclear shape modes therefore also captured variation between HCs and SCs, although this information was represented by multiple shape modes.

Unsupervised Clustering of Neuromast Cells in Shape Space Identifies Groups of Cells With Similar Shapes and Locations

We next determined whether shape features could be used to assign neuromast cells to biologically meaningful categories beyond HCs and SCs. We performed unsupervised clustering on the first 8 cell shape PCs using the Phenograph package and Leiden community detection algorithm (Levine et al., 2015; Traag et al, 2019). Five identified clusters were represented in two dimensions by a uniform manifold approximation and projection (UMAP; McInnes et al., 2018; Figure 4A) where each cell is a point colored by cluster label. The fraction of cells within each cluster was proportional across all individual neuromasts (Supplemental Figure 6D) suggesting that clustering is influenced by shared shape parameters and not due to differences between neuromasts. To quantify the relationships between clusters, we performed partition-based graph abstraction (PAGA; Wolf et al., 2019). Relative strength of association is represented by line thickness (Figure 4B). PAGA analysis revealed greater similarities between clusters 0/2, 1/2/3, and 2/4.

To visualize the characteristic shapes of cells within each cluster, we identified the cell closest to the centroid of each cluster in eight-dimensional PC space as a “representative cell” (Figure 4C). Cells in clusters 0 tended to be shorter with greater depth, while cells in cluster 3 tended to be tall and wide. Cells in clusters 1 and 2 appeared to have intermediate shapes. Cells in cluster 4 tended to have flask shapes reminiscent of HCs; quantifying manually annotated HCs in each cluster confirmed that most HCs were assigned to cluster 4 (Supplemental Figure 6E) and cluster 4 was primarily composed of HCs (Supplemental Figure 6F).

Cells from a given cluster tended to be found at distinct radial distances from the neuromast center (Figure 4D,E), and we used this order to assign cluster number labels. However, we also found that some clusters were distributed in distinct quadrants of the neuromast (Figure 4D,F). This difference in distribution was particularly apparent for cluster 1, where cells were more concentrated within the dorsoventral quadrants (Figure 4F, represented by adjusted angle 90 and -90), and cluster 2, where cells were more often found within the anteroposterior quadrants (Figure 4F, represented by adjusted angle 0). These results suggest that grouping neuromast cells purely using shape descriptors, without explicitly including information about cell type or identity, recapitulates known aspects of neuromast organization (e.g. concentric and radially symmetric cell populations).

We also performed unsupervised clustering on the first 4 nucleus shape PCs. Similar to cell shape clusters, we found distinct nucleus shape clusters (Figure 5A) whose relative relationships could be measured through PAGA analysis (Figure 5B). As for cells, we identified representative nuclei for each cluster (Figure 5C). Clusters 4 and 5 were primarily composed of HCs, and cluster 1 was enriched for dorsoventral cells (Figure 5D). Nucleus shape clusters exhibited distinct spatial locations within the neuromast (Figure 5D,E,F). However, nucleus shape clusters did not appear to distinguish as well between SC subpopulations compared to cell shape clusters. Since the nucleus shape dataset contains cells that were excluded from cell shape analysis, we looked at the distribution of excluded cells across nucleus shape clusters (as a way to understand whether a particular class of cells had been excluded). We found excluded cells in all 5 nucleus shape clusters, suggesting that excluded cells can have a full range of nucleus shape phenotypes (Supplemental Figure 1E). Based on these results, we decided to perform the remaining analysis using only cell shape clusters.

Cell Fate Markers of SC Subpopulations Are Associated With Distinct Cell Shape Phenotypes

SC subpopulations have characteristic transcriptional profiles and behaviors during regeneration: DV cells act as HC progenitors while peripheral cells can replenish DV cells depleted by HC regeneration (Lush et al., 2019; Thomas and Raible, 2019). We therefore examined neuromasts from knock-in lines that mark these distinct SC subpopulations. The *sost:NLS-Eos* line marks cells in the DV compartment of the neuromast (Figure 6A-C). We measured the mean intensities of Eos fluorescence in each cell, normalizing to z-score for each individual neuromast. We then classified cells as Eos+ if they had fluorescence with z-score greater than 1 (Figure 6B, Supplemental Figure 6A). An example neuromast with labeled positive cells is shown in Figure 6B. As expected, *sost:NLS-Eos*+ cells were found in DV quadrants (Figure 6B,C). We next asked whether *sost:NLS-Eos*+ cells were differentially distributed across clusters (Figure 6E,F). Most *sost:NLS-Eos*+ cells were found in cluster 1 (Figure 6E). 33% of cluster 1 cells were s*ost:NLS-Eos*+, the highest proportion of any cluster (Figure 6F). Some *sost:NLS-Eos*+ cells were also found in clusters 2 and 3,making up 13% and 17% of clusters 2 and 3, respectively (Figure 6F).

The *sfrp1a:NLS-Eos* line marks peripheral SCs (Figure 6G-I). We used the same threshold method to classify cells as Eos+ (Figure 6H, Supplemental Figure 6B). Eos+ cells were found in the neuromast periphery (Figure 6H,I). Most *sfrp1a:NLS-Eos*+ cells were found within cluster 0, making up about half of the cells in this cluster (Figure 6K,L). Some *sfrp1a:NLS-Eos*+ cells were also found in clusters 1 and 3, making up 13% and 29% of clusters 1 and 3, respectively. Clusters showed no differences in distribution among individual neuromasts across wildtype, *sost:NLS-Eos*, and *sfrp1a:NLS-Eos* lines (Supplemental Figure 6D). Taken together, these results demonstrate that cells with distinct identities and localization can be classified by distinct cell shape characteristics.

atoh1a Mutants Exhibit Changes in Cell Shape Across Cell Populations

We next sought to test what happens to cell shape distributions when a single cell type is absent. The transcription factor *atoh1a* is required for HC fate specification in zebrafish lateral line neuromasts (Milimaki et al., 2007; Itoh and Chitnis, 2001). We expected that the loss of the HC shape phenotype would be readily observable in *atoh1a* mutants. However, we wondered whether SCs would maintain the same shapes seen in wildtype fish or if their shapes would also be altered in *atoh1a* mutants. In the *atoh1a:mRuby* line, the mRuby fluorophore was inserted into the coding sequence of the *atoh1a* gene, disrupting its function but acting as a readout for *atoh1a* promoter activity. This knock-in line enables comparisons not only between wildtype, mutant, and heterozygous cell populations as a whole, but also between mRuby+ and mRuby- populations in each genotype.

In heterozygous fish, mRuby+ cells are localized to the center of the neuromast (Figure 7A; Supplemental Figure 6E). Cells classified as mRuby+ had fluorescence with z-score greater than 1 (Figure 7B, Supplemental Figure 6C). Overall, the proportions of cells in each cluster were similar to those of wildtype fish (Supplemental Figure 6D,G). However, mRuby+ cells were differentially proportioned across clusters and highly enriched in cluster 4 (Figure 7C,D), which contains a high proportion of hair cells (Supplemental Figure 6H,I).

In homozygous mutant fish, mRuby+ cells were still found in the center of the neuromast even though they did not differentiate into HCs (Figure 7E,F; Supplemental Figure 6E). There was no statistically significant difference in the proportion of mRuby+ cells between heterozygous and homozygous animals ($X^2 = 1.8$, df = 1, p = 0.18; Supplemental Figure 6F). mRuby+ cells in homozygotes were differentially distributed across clusters (Figure 7G,H; Supplemental Figure 6D,G). Relative to wildtype and heterozygotes, mutant fish had far fewer cells in cluster 4 (Figure 7H; Supplemental Figure 6D,G). By contrast, the number of mRuby+ cells in clusters 2 and 3 were expanded in mutants. For mRuby- cells, cluster 2 was expanded in mutants relative to heterozygotes (Figure 7D,H). These results suggest that loss of *atoh1a* function is associated with a loss of the HC shape phenotype, an effect mainly seen in mRuby+ cells. The loss of HCs did not result in the appearance of a new cell shape cluster but rather assignment of cells into existing shape clusters, mainly cluster 2. These results demonstrate that cell shape analysis can be used to query changes in phenotype in contexts where development has been perturbed. Additionally, these results show that in neuromasts some aspects of cell shape are genetically determined, rather than being solely due to other factors (such as position).

Shape Modes Can Be Used to Classify HCs

We next examined the distribution of CSM1 scores from *atoh1a* heterozygous and mutant fish. We found that the distribution of CSM1 scores in heterozygotes resembled that of wildtype fish, with HCs associated with positive scores (Figure 7I, top and middle). By contrast, the distribution of CSM1 scores in homozygous mutants was no longer bimodal (Figure 7I, bottom) and was skewed negative and corresponded to the distribution of SCs in wildtype and heterozygous animals. CSM1 therefore captured the variation in shape between the two broadly defined neuromast cell populations (HCs and SCs) and was predictably perturbed in mutants lacking HCs.

The distinction between HCs and SCs in CSM1 scores suggested that cell shape modes may contain sufficient information for classification. We trained a logistic regression classifier to predict HC identity given the first eight cell shape mode scores (CSM1-8). The Receiver Operating Characteristic (ROC) curve indicated robust classifier performance with Area Under the Curve (AUC) of 0.96 (Figure 7J). Despite the imbalance of classes in the dataset (77% SCs and 23% HCs), the classifier correctly labeled most (78%, 100/128) of the HCs in the test set, while few SCs (5%, 22/422) were incorrectly labeled as HCs (Figure 7K). We also asked if HCs could be classified using nucleus shape features, training a classifier using the first four nucleus shape modes (NSM1-4). The nuclear shape classifier performed similarly well to one using cell shape features (AUC score = .0.96; data not shown). Therefore, nuclear shape features can also be used to classify HCs from SCs.

To visualize the shape characteristics associated with HC identity, we created a 3D reconstruction of the idealized mean HC and SC in the dataset by calculating the mean CSM1-8 scores for each cell type and deriving the corresponding SHE coefficients (Figure 7L). As expected, the “average” HC has a flask-like morphology, while the “average” SC has a concave shape. These results demonstrate the utility of cell shape features both for describing and classifying cell types.

Discussion

Cells transition through states as they grow, divide, and differentiate. Measuring gene transcription has proven to be a powerful method for understanding and classifying cell types. However, cells can change in ways that are not necessarily reflected in RNA transcript levels (Marklein et al., 2016; Gerbin et al., 2021). Cell shape analysis offers another avenue to quantitatively characterize and classify cells. Methods to encode cell shape in an unbiased, high dimensional representation have been employed with great success to cells in culture (Pincus et al., 2007; Keren et al., 2008; Viana et al., 2023). However, most studies of cell shape in living, developing animals have relied on 2D projections of cells or representative 3D geometric features, which may not capture many aspects of cell shape in these tissues. In this study, we use SH and PCA to encode cell shape information in a more complete, unbiased fashion, allowing for data-driven analysis of the relationships between cell shape and fate.

A primary goal of our study was to understand the relationship between neuromast cell shape and established indicators of cell identity. If cell shape and markers for neuromast cell fate were tightly correlated, then it might be possible to classify cell types using shape features. Our results suggest that this relationship was true of HCs, since we were able to build a logistic classifier for HCs using our shape modes. When we performed unsupervised clustering of cells in shape space, there was a high degree of overlap between HC identity and one of the clusters (HCs comprised 79% of cluster 4). We believe similar methods could be used to build classifiers for cell types in other developmental contexts, potentially allowing for cell fate to be predicted from image-based data without expression of cell fate markers.

The association between cell shape and SC subtypes was more nuanced. In our clustering analysis, DV cells defined by *sost:NLS-Eos* and peripheral cells defined by *sfrp1a:NLS-Eos* were enriched in distinct clusters. However, these clusters also contained substantial proportions of cells not expressing the marker. Given that SCs act as multipotent progenitors (Thomas et al., 2019; Romero-Carvajal et al., 2015; Viader-Llargues et al., 2018) we speculate that perhaps the more continuous distribution of SC shapes may reflect this plasticity. In contrast, HCs are a terminally differentiated cell type, which may be represented in their more distinct cell shape phenotype. A caveat is that our cutoff for considering a cell positive for a marker was relatively strict. In addition, our transgenic reporters may not always reflect the cells most recently expressing the transgene due to the high stability of the NLS-Eos protein (Cruz et al., 2015). Future studies using other methods to label SC subpopulations with genetic markers, such as RNA-FISH, could further elucidate the connection between gene expression and cell shape in neuromasts. More broadly, cell shape analysis could be integrated with other cellular properties, such as RNA transcript expression, protein expression and organelle location.

We were also interested to see whether nuclear shape could be used to classify cell types. Because nuclei are often easier to label and segment, nuclear shape representations may be more readily attainable in some developmental biology contexts. We found that nuclear shape could be used to reliably classify HCs from SCs. However, compared to cell shape features, nuclear shape features did not appear to distinguish as strongly between the SC subpopulations we tested in this study. However, nucleus shape may still be informative in other ways – we observed spatial patterns in nucleus shape within neuromasts, which may represent differences in cell type or state that were not tested in this study. Analysis of nucleus shape may also be more amenable to other tissues where cell membranes may not be easily resolved, allow more rapid collections of larger datasets, or allow analysis of cell types whose cell shapes are not amenable to SH analysis. More broadly, we imagine that cell shape, nuclear shape, or a combination of the two may be useful as predictors of cell fate in different tissues.

In many developmental contexts, it remains unclear how genetic factors determine cell shape. We observed that loss of function in *atoh1a* results in the large reduction of the cluster associated with HCs and concurrent increases in existing clusters, but does not lead to generation of a new shape cluster. Similarly, during an RNAi screen for morphological complexity in *Drosophila* Kc cells, RNAi knockdown typically did not generate new shapes but instead changed distributions of pre-existing, wildtype shape states (Yin et al., 2013). It would be illuminating to explore other genes and signaling pathways that may pattern cell shape within neuromasts. For example, Fgf signaling has been shown to be required for apical constriction and rosette formation in the lateral line primordium (Harding and Nechiporuk, 2012; Ernst et al., 2012). Applying cell shape analysis to query other genetic and pharmacological perturbations could elucidate how cell shape in neuromasts is regulated by different signaling pathways.

Previous work has proposed that HC death might induce cell shape changes and tissue deformations that trigger SCs to proliferate; this idea originates from observations that (1) mechanical strain and cellular spreading induce proliferation in some contexts (Chen et al., 1997; Watt et al., 1988; Benham-Pyle et al., 2015) and (2) SC shape change during HC death is positively correlated with an organism's HC regeneration capability (Burns et al., 2013; 2014). Mechanical cues and cell density might also act as a brake on proliferation during regeneration, similar to what has been observed during development in the murine utricle (Gnedeva et al., 2017). Quantifying cell shape change during HC death or targeted ablation of neuromast cells could query the role of cell shape in regulating proliferation and regeneration. Of particular interest is Yap, a transcriptional coactivator that regulates cell proliferation downstream of mechanical cues (Piccolo et al., 2014) known to play a role in regulating proliferation during development and regeneration of HCs in mammals (Rudolf et al., 2020; Gnedeva et al., 2020; Kastan et al., 2021). Coupling cell shape analysis with reporters for Yap activity could shed light on whether cell shape change acts upstream of Yap-induced cell proliferation. Cell shape analysis could also be combined with techniques to measure and perturb mechanical forces.

Previously, Hartmann and colleagues (2020) analyzed 3D cell morphology in the migrating posterior lateral line primordium using a novel method (ISLA-CBE) based on point cloud morphometry. A strength of ISLA-CBE is that rotational variation can be minimized using a pairwise distance transformation. In addition, ISLA-CBE is not limited only to cells that can be accurately spherically parameterized. However, ISLA-CBE does not easily allow cell shapes to be reconstructed from any point in feature space, an advantage of our method. We anticipate that both methods will be useful in the future, as the ideal method of cell shape encoding will likely depend on the context and goals of the study. Although the method of encoding cell shape was different, we observed some commonalities with this study. Our observations suggested that HCs represent a discrete morphological state from SCs, while SCs appeared to vary more continuously in shape. Similarly, Hartmann and colleagues (2020) observed a continuous spectrum of shapes, without distinct clusters, in the lateral line primordium. Together, these results suggest a model whereby initially, cells in the lateral line vary continuously in shape, but as HCs differentiate, they emerge as a discrete shape. How early these patterns develop and whether they change over time or remain static during neuromast deposition and maturity, could be addressed in future studies.

Although our study looked only at homeostatic, mature neuromasts at one time point, we believe the cell shape representation used here could be valuable for integrating information across multiple spatial and temporal scales. Using time lapse imaging and cell tracking, it would be possible to study dynamic cell shape changes during HC development, death, and regeneration. We envision that cell shape states could be ordered along trajectories that represent developmental time, analogous to pseudotime analysis employed in single cell RNA sequencing studies. In addition, shape analysis can be applied to understand the morphologies of whole tissues and organs (Dalmasso et al., 2022). Combining pseudotime analysis with tissue-scale measurements can elucidate the relationship between individual cell shape change and tissue/organ morphogenesis, as has been demonstrated in a study of amphioxus notochord development (Andrews et al., 2021). Quantifying how these relationships evolve as the animal grows could shed light into how individual cell shape changes and movements shape organs, as well as how, conversely, the shape of organs constrains individual cell shape. Comparative cell shape studies could also be done between different sensory epithelia in the same organism or even between different species or ages, which could inform evolutionary relationships.

Limitations of the Study

Our machine-learning based method to segment cells may well be specific to the specific labeling conditions of our transgenic line and our live imaging methods. In addition, the method was only semi-automated and required some manual curation and correction. We may see improvement if we retrain the model with this expanded, corrected dataset as the ground truth. However additional adjustments would likely be needed for cells labeled with other reagents. As more sophisticated methods for automated segmentation become available, e.g. Cellpose (Pachitariu and Stringer, 2022), alternative pipelines may require less supervision. We also note that the resulting cell and nuclear shapes will likely be influenced by the specific characteristics of fluorescent labels and segmentation algorithms used, which will in turn influence subsequent SH analysis. We therefore constrained our analysis to the same nuclear and membrane label across different fish lines.

In the SH-based approach used here, all points on the surface of the cell must be able to be mapped onto a sphere to generate accurate reconstructions. However, some of our cells are sufficiently concave that this is not possible, since a line drawn from the center to the outside of the surface would intersect the surface more than once. Since this problem likely extends to other cell types beyond neuromasts, there is a need for other 3D cell shape analysis methods that can better represent complex, concave shapes. Development and application of alternative methods for spherical parameterization and SH expansion offer one potential solution to this problem. For example, Dalmasso and colleagues (2022) represented cell shapes as signed distances from concentric spheres, which enabled them to represent concave structures of mouse limbs. Another method using 4D hyperSH enables multiple disjointed objects to be parameterized together, and also enables representation of shapes with gaps or holes (Pasha Hosseinbor et al., 2015). There are also alternatives to SH for data-driven 3D shape representations, including deep learning approaches such as autoencoders (De Vries et al., 2022). With the rate of exciting developments in the field, we expect that many methods will be available to study cell shape in a variety of contexts in the future.

Materials and methods

Fish lines

Experiments were conducted on 5 dpf larval zebrafish (*Danio rerio*). The following zebrafish lines were used for these studies: *Tg(-8.0cldnb:LY-EGFP)*^{zf106Tg} [RRID:ZDB-ALT-060919-2] (Haas and Gilmour, 2006), *sost*^{w215Tg} [RRID: ZDB-ALT-190909-10] , *sfrp1a*^{w217Tg} [RRID:ZDB-ALT190909-6] (Thomas and Raible, 2019), *atoh1a*^{w271Tg} (this paper). Animals were used at a stage before sex is determined.

Generation of *atoh1a*^{w271Tg}

The *atoh1a*^{w271Tg} line was generated by CRISPR-Cas9 knock-in of the mRuby fluorophore into the coding sequence of the endogenous *atoh1a* locus, as previously published (Thomas and Raible, 2019; Kimura et al., 2014). Homozygotes exhibit phenotypes consistent with *atoh1a* loss of function, such as hearing and vestibular defects, and are viable until approximately 10 days post fertilization (dpf).

Fish handling

Prior to experiments, embryos and larvae were raised in E3 embryo medium (14.97 mM NaCl, 500 µM KCL, 42 µM $NaHPO_4$, 150 µM KH_2PO_4, 1 mM $CaCl_2$dehydrate, 1 mM $MgSO_4$, 0.714 mM $NaHCO_3$, pH 7.2) at 28.5°C. Zebrafish larvae were fed rotifers daily from 4 dpf onwards. Experiments used zebrafish at 5 dpf, a stage before sex is determined. Zebrafish experiments and husbandry followed standard protocols in accordance with University of Washington Institutional Animal Care and Use Committee guidelines.

DRAQ5 staining

On the day of imaging, DRAQ5 dye (Thermo-Fisher) was diluted in E3 embryo medium to a working concentration of 5 µM. 5 dpf larval zebrafish were incubated in this solution for approximately 1 hour, then prepared as described in "Preparation of larvae for live imaging." We observed best results when the fish were incubated in DRAQ5 no longer than 1 hour and immediately prepared for imaging afterward.

Photoconversion of NLS-Eos

For experiments done with *sost:NLS-Eos* or *sfrp1a:NLS-Eos* lines, larvae were photoconverted using an iLumen 8 UV flashlight (purchased from Amazon) for 15-20 minutes. Photoconversion was done by placing fish in a 60 × 15 mm petri dish, removing the petri dish lid, and placing the UV flashlight directly over the dish within a box lined with aluminum foil.

Preparation of larvae for live imaging

Fish were transferred to E3 embryo medium containing approximately 1.5 mM MESAB to be anesthetized. Once larvae were suitably anesthetized, they were mounted on slides with bridged coverslips in 1.2% low melting point agarose. Fish were mounted in a standard orientation of anterior facing left and dorsal up.

Confocal imaging

Imaging was performed using a Zeiss LSM 880 with Airyscan (Huff, 2015) equipped with a Zeiss C-Apochromat 40x/1.2W numerical water objective. Imaging was performed at room temperature (approximately 25°C). For each fish, several anterior neuromasts were imaged. Neuromasts were chosen from a list based on which had the most ideal orientations in the fish being imaged (e.g. those neuromasts closest to the coverslip with apical-basal axes roughly parallel to the axial plane/Z-axis of imaging). 633 nm, 561 nm, and 488 nm lasers were used for excitation. Gain was adjusted for different transgenes but kept constant across experiments with the same transgene. Images were collected with physical voxel size of 0.05 x 0.05 x 0.22 μm^3. All images were captured using the Zen Black acquisition software (Zeiss). Minimal cell death was observed during imaging; any neuromasts exhibiting significant amounts of cell death were excluded from further analysis.

Image preprocessing

Using Zen Blue, Airyscan Processing was applied to all raw images using the default settings for pixel reassignment and deconvolution. Next, a custom Python script using the aicsimageio package (Brown et al., 2021) was used to remove any unneeded channels and export the CZI files to TIFFs. The original CZI files were kept to ensure access to the original acquisition metadata. To correct for xy drift between z-slices, StackReg was used, either through the MultiStackReg FIJI plugin or a custom script using the pystackreg Python port of StackReg (Thevenaz et al., 1998). Using the “rigid body” method and “previous” setting for the reference, transformation matrices were calculated for the membrane channel, saved, and applied to all channels in the image. Images were manually inspected following stack registration to ensure appropriate drift correction.

Occasionally, imaging acquisition errors or z drift caused either duplication of z slices within the stack or the need to acquire additional z-slices to cover the entire neuromast. In cases where this could be easily manually corrected, duplicated z-slices were deleted, or two zstacks were concatenated together to cover the entire neuromast. These manual corrections were done using FIJI (Schindelin et al 2012) prior to stack registration.

Training of machine learning models for detection of nuclei and membranes

Models to detect neuromast nuclei and membranes were trained using the iterative deep learning workflow of the Allen Cell and Structure Segmenter (Chen et al., 2018). For the membrane model, preliminary segmentations of cell boundaries were initially generated using the lamin B1 workflow from the Segmenter. The "Curator" tool from the Segmenter was used to mask out areas of the images that were not segmented accurately. The original membrane model was trained on a dataset of 14 neuromasts labeled with *Tg(-8.0cldnb:LY-EGFP)*. The membrane model predictions were thresholded and used as input for an interactive distance transform watershed segmentation workflow. Upon generating the label images, the find_boundaries function from scikit-image (van der Walt et al., 2014) was used to generate watershed lines, which was used as input to train a second membrane model with improved performance. This process was repeated once more with a new dataset of 9 *Tg(-8.0cldnb:LYEGFP)* neuromasts stained with DRAQ5, except labeled nuclei were used to generate seeds for the watershed instead of the distance transform. The results were used to train the third membrane model, which was used for instance segmentation (delineation of individual objects) of all the images in this study.

To generate the nucleus mask model, the "H2B coarse" model provided by AICS[22] was used to generate preliminary nucleus predictions. The predictions were thresholded, and the membrane intensity image (or membrane boundary predictions) were used to "split" the nuclei (as described in "Instance segmentation"). Similar to the refinement for the membrane model, a distance transform watershed workflow was applied to thresholded nuclei predictions to generate labeled nuclei, which were converted to binary masks and used to train the first nucleus mask model. This process was repeated (starting from the model 1 predictions instead of the H2B coarse model predictions) to generate a second nucleus mask model with improved performance, which was used for all the images in this study.

Instance segmentation

Instance segmentation of nuclei was done in a semi-automated, interactive fashion using a custom Python script. The Python packages napari (Soferoniew et al., 2022) and magicgui (Lambert et al., 2023) were used to create a graphical user interface for this step, enabling parameter tuning, interactivity, and manual annotations when needed.

First, the nucleus mask predictions and cell boundary predictions were binarized using thresholds of 0.6 and 0.4, respectively. These thresholds were selected a balance between capturing intensity signal corresponding to the structures of interest and minimizing artifacts due to noise. The binarized membrane predictions were used to “split” falsely merged nuclei in the binarized nucleus mask predictions, using one of two methods. For experiments not involving NLS-Eos transgenes, the nucleus mask splitting was done fully automatically (e.g. wherever the membrane mask value was 1, the corresponding pixels in the nucleus mask were set to 0). For experiments involving NLS-Eos transgenes, incomplete photoconversion and/or ongoing transgene expression can cause some signal in the nuclei in the green channel, where the membranes are also imaged. This overlap can create artifacts when the cell boundary model is applied. To circumvent the issue, an alternative “interactive” nucleus splitting strategy was devised. For interactive nucleus splitting, regions of interest (ROIs) were manually drawn sliceby-slice in areas where splitting was desired using the shapes tool in napari. Then, only these regions within the membrane mask were considered during the splitting process.

Following nucleus splitting, individual nuclei were labeled using distance transform watershed. First, the Euclidean distance transform of the nucleus mask was calculated and smoothed using a Gaussian kernel. Next, the local maxima of the smoothed distance transform were calculated using the peak_local_max function from scikit-image. The local maxima were then used as markers for a watershed segmentation using the inverted distance transform as the input. The sigma parameter for the Gaussian smoothing of the distance transform and the minimum distance parameter for the peak_local_max function were adjusted interactively to yield the best results for each image. Remaining segmentation errors were corrected by manual annotation using a Wacom Cintiq 16 drawing tablet and built-in labels layer tools in napari.

To segment cells, the segmented nuclei were used as markers for a watershed segmentation (Beucher and Lantuejoul, 1979) with the cell boundary predictions (i.e. the raw, unthresholded output of the membrane model inference) as input. The resulting preliminary cell labels were post-processed to remove small objects and manually corrected using a Wacom Cintiq 16 drawing tablet and built-in labels layer tools in napari. After segmentation, each field of view was manually inspected, and some cells were excluded for quality control (e.g. due to having poor signal or extending beyond the edge of the field of view). 164 cells were manually excluded in this way.

Cell dataset preparation

Prior to running the cell variance analysis pipeline (Viana et al., 2023) individual cells from the raw and labeled whole neuromast images were cropped, subsampled in z to obtain isotropic voxel size, and saved as individual images. A custom Python script was used to prepare cropped single cells and a manifest.csv file compatible with the pipeline.

Cell and nucleus alignment

Cells were aligned using a series of three rotations. The first rotation was done with reference to the neuromast organ. The neuromast centroid was calculated from a mask representing the whole neuromast (created from the cell segmentation channel). For each cell, the relative position of the cell centroid to the neuromast centroid was used to calculate the angle between the cell centroid and the x-axis. The corresponding cropped image of the cell was then rotated around the z-axis by this angle to minimize variation originating from the rosette structure of the neuromast.

The second and third rotations were done with reference to the cropped cell only. For the second rotation, the cropped image of the cell was rotated such that the long axis of the cell projected in the xz-plane was aligned to the z-axis. For the second rotation, the cropped image of the cell was rotated such that the long axis of the cell projected in the yz-plane was aligned to the z-axis, unless this angle would be greater than 45 degrees. In that case, the yz long axis was aligned to the y-axis instead. This adjustment prevented cells from being incorrectly rotated by 90 degrees if their long axis was in the horizontal plane.

To generate the nucleus shape space, the nuclei were aligned separately from the cells. The procedure was the same as for the cells, except for the second rotation which also was constrained to be no greater than 45 degrees. If the angle would be greater than 45 degrees, the long axis of the nuclei in the xz-plane was aligned to the x-axis instead of the z-axis.

Spherical harmonics parameterization

Following cell dataset preparation and alignment, SH coefficients for each cell and nucleus were calculated using the cvapipe_analysis package (Viana et al., 2023) First, the “loaddata” step was run with the prepared manifest.csv, then the “computefeatures” step was run. To determine the degree of expansion (L_{max}) for SH parameterization, we plotted several metrics of reconstruction fidelity for different values of L_{max} using a subset of the data. As L_{max} increases, shapes can be reconstructed with greater detail, but higher L_{max} values incur additional computational costs and exhibit diminishing returns in reconstruction fidelity. We selected an L_{max} of 32 as a value around which reconstruction fidelity plateaued, which yielded 2178 coefficients. The directed Hausdorff distance (Huttenlocher et al., 1993), which represents the furthest distance between the closest points on the original and reconstructed meshes, was used as a measure of reconstruction error. A Gaussian mixture model was used to identify cells belonging to the second peak in the reconstruction error distribution, which were subsequently excluded from cell shape analysis.

Principal components analysis and unsupervised clustering

Principal components analysis (PCA) was performed using the Python library scikit-learn (Pendregosa et al., 2011). To create “shape modes,” PCs were z-scored (i.e. the values for each PC were divided by their respective standard deviation). Visualizations of cells in shape space were generated by applying the inverse PCA transform to generate corresponding SHE coefficients, which were then used to generate 3D reconstructions using the aics-shparam package.[10] Reconstructions were visualized in 3D using ParaView (Ahrens et al., 2005).

Clustering was performed using the Python package Phenograph (Levine et al., 2015) with the Leiden algorithm for community detection (Traag et al., 2019). For analysis of cell shape, the first eight PCs were used for clustering; for nuclear shape, the first four PCs were used. These PCs explained 70% and 76% of total variance, respectively. The Uniform Manifold Approximation and Projection (UMAP) implementation from the Python package umap-learn was used to project the data in two dimensions for visualization (McInnes et al., 2018) Partition based graph abstraction (PAGA) was performed using the scanpy Python package (Wolf et al., 2018; 2019) PAGA was used to initialize the layout of UMAPs.

Quantification of fluorescence intensity of transgenes

For reporter lines expressing nuclear localized Eos, *sost:NLS-Eos* and *sfrp1a:NLS-Eos*, the mean intensity within each cell was calculated using the segmented nuclei as masks. For *atoh1a:mRuby*, which has cytoplasmic mRuby localization, the mean intensity within each cell was calculated using the segmented cell as masks. Because intensities of fluorescent proteins from targeted insertions can vary between individual neuromasts and fish, we z-scored the mean intensities for each neuromast and used a threshold of 1 to determine cells “positive” for a marker. In other words, the mean intensity of fluorescent protein within a cell needed to be at least one standard deviation above the mean for that neuromast to be classified as “positive” for that protein.

Measurement of cell location and shape features

Locations of neuromast cells were estimated using the x and y coordinates of the cell segmentation centroids. The neuromast centroid was calculated by using the cell segmentation channel to generate a mask representing the whole neuromast. For each cell, the distance between the cell centroid and neuromast centroid in the xy plane was calculated. This distance was normalized to that of the cell with the greatest distance from the neuromast centroid to account for differences in neuromast size (Vlader-Llargues et al., 2018).

We used the first alignment rotation angle to approximate the angular location of each neuromast cell. This angle represents the angle between the neuromast x-axis (which extends from the neuromast centroid) and a vector drawn from the cell centroid to the neuromast centroid in the xy-plane (Figure 2B). In our images, the neuromast x-axis is approximately parallel to the anterior-posterior body axis of the fish. Lateral line neuromasts can be polarized in one of two ways relative to the anteroposterior body axis: parallel or perpendicular (LopezSchier et al., 2004). Because our dataset contained neuromasts of both types, we standardized the orientations of the neuromasts by adjusting the angles of perpendicular neuromasts using the following formula:

$$if\, o \leq 90: a = o + 90$$

$$if\, o > 90: a = 360 - (o + 90)$$

where *o* is the original angle, and *a* is the adjusted angle. This is equivalent to rotating a perpendicular neuromast 90 degrees counterclockwise such that its dorsoventral axis is aligned with the anteroposterior body axis. The decision to rotate counterclockwise was based on the fact that Emx2 is expressed in the anterior region of parallel neuromasts and the dorsal region of perpendicular neuromasts (Jiang et al., 2017). The radial distances and angular locations of each cell were combined to plot the locations of neuromast cells on polar coordinates.

Other shape measurements, such as volume, height, etc. were obtained as outputs of the cvapipe_analysis computefeatures step (Viana et al., 2023)

Logistic regression classification

Logistic regression was performed using the Python package scikit-learn (Pedregosa et al., 2011). The data were split into an 80/20 ratio of training data to test data. The test and train subsets were split such that both had similar proportions of HCs. The logistic regression model was trained to classify HCs on the train subset, using one or more shape modes as predictor variables. The trained model was then applied to the testing subset to generate the ROC curves and confusion matrices.

Statistical Analysis

Chi-square analysis was used to distinguish distribution differences among groups, using the Python package SciPy (Virtanen et al., 2020).

Acknowledgments

The authors thank David White and the UW Fish Facility staff for animal care and maintenance. We thank Julie Theriot and members of her lab for insightful discussion and feedback throughout the course of this study. We also thank Celeste Berg and Cecilia Moens for critical reading of the manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by grant awards R21 DC015110 and R21 DC019948 from the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health, Hearing Health Foundation, Hamilton and Mildred Kellogg Trust, and the Whitcraft Family Gift to DWR. MNH was supported by grant awards T32 DC005361 and F31 DC019028 from NIDCD.

Data and Code Availability

Image data is deposited and publicly available at the BioImage Archive (https://www.ebi.ac.uk/biostudies/studies/S-BIAD886). All original code is publicly available at the date of publication at https://github.com/raible-lab/neuromast3d.

References

Ahrens, J., Geveci, B., and Law, C. (2005). ParaView: An End-User Tool for Large-Data Visualization. In *Visualization Handbook* (Elsevier), pp. 717–731.

Andrews, T.G.R., Pönisch, W., Paluch, E.K., Steventon, B.J., and Benito-Gutierrez, E. (2021). Single-cell morphometrics reveals ancestral principles of notochord development. *Development* **148**, dev199430.

Benham-Pyle, B.W., Pruitt, B.L., and Nelson, W.J. (2015). Mechanical strain induces Ecadherin–dependent Yap1 and β-catenin activation to drive cell cycle entry. **Science 348**, 1024–1027.

Bermingham, N.A., Hassan, B.A., Price, S.D., Vollrath, M.A., Ben-Arie, N., Eatock, R.A., Bellen, H.J., Lysakowski, A., and Zoghbi, H.Y. (1999). Math1: An Essential Gene for the Generation of Inner Ear Hair Cells. *Science* **284**, 1837–1841.

Beucher, S., and Lantuéjoul, C. (1979). Use of Watersheds in Contour Detection. In *International workshop on image processing, real-time edge and motion detection*.

Brown, E.M., Toloudis, D., Sherman, J., Swain-Bowden, M., Lambert, T., and AICSImageIO Contributors (2021). AICSImageIO: Image Reading, Metadata Conversion, and Image Writing for Microscopy Images in Pure Python. [Computer software]. GitHub. https://github.com/AllenCellModeling/aicsimageio

Burns, J.C., Collado, M.S., Oliver, E.R., and Corwin, J.T. (2013). Specializations of intercellular junctions are associated with the presence and absence of hair cell regeneration in ears from six vertebrate classes. *J. Comp. Neurol.* **521**, 1430–1448.

Burns, J.C., and Corwin, J.T. (2014). Responses to Cell Loss Become Restricted as the Supporting Cells in Mammalian Vestibular Organs Grow Thick Junctional Actin Bands That Develop High Stability. *J. Neurosci.* **34**, 1998–2011.

Chan, C.J., Heisenberg, C.-P., and Hiiragi, T. (2017). Coordination of Morphogenesis and CellFate Specification in Development. *Curr. Biol.* **27**, R1024–R1035.

Chen, C.S., Mrksich, M., Huang, S., Whitesides, G.M., and Ingber, D.E. (1997). Geometric Control of Cell Life and Death. *Science* **276**, 1425–1428.

Chen, J., Ding, L., Viana, M.P., Lee, H., Sluezwski, M.F., Morris, B., Hendershott, M.C., Yang, R., Mueller, I.A., and Rafelski, S.M. (2018). The Allen Cell and Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images. *bioRxiv* 491035; doi: https://doi.org/10.1101/491035

Cruz, I.A., Kappedal, R., Mackenzie, S.M., Hailey, D.W., Hoffman, T.L., Schilling, T.F., and Raible, D.W. (2015). Robust regeneration of adult zebrafish lateral line hair cells reflects continued precursor pool maintenance. *Dev. Biol.* **402**, 229–238.

Dalmasso, G., Musy, M., Niksic, M., Robert-Moreno, A., Badía-Careaga, C., Sanz-Ezquerro, J.J., and Sharpe, J. (2022). 4D reconstruction of murine developmental trajectories using spherical harmonics. *Dev. Cell* **57**, 2140-2150.e5.

De Vries, M., Dent, L., Curry, N., Rowe-Brown, L., Bousgouni, V., Tyson, A., Dunsby, C., and Bakal, C. (2022). 3D single-cell shape analysis using geometric deep learning bioRxiv 2022.06.17.496550; doi: https://doi.org/10.1101/2022.06.17.496550

Ernst S, Liu K, Agarwala S, Moratscheck N, Avci ME, Dalle Nogare D, Chitnis AB, Ronneberger O, Lecaudey V. (2012). Shroom3 is required downstream of FGF signalling to mediate proneuromast assembly in zebrafish. *Development*. **139**:4571-81.

Gerig, G., Styner, M., Jones, D., Weinberger, D., and Lieberman, J. (2001). Shape analysis of brain ventricles using SPHARM. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis* (MMBIA 2001), pp. 171–178.

Gnedeva, K., Jacobo, A., Salvi, J.D., Petelski, A.A., and Hudspeth, A.J. (2017). Elastic force restricts growth of the murine utricle. *eLife* **6**, e25681.

Gnedeva, K., Wang, X., McGovern, M.M., Barton, M., Tao, L., Trecek, T., Monroe, T.O., Llamas, J., Makmura, W., Martin, J.F., et al. (2020). Organ of Corti size is governed by Yap/Teadmediated progenitor self-renewal. *Proc. Natl. Acad. Sci.* **117**, 13552–13561.

Gerbin KA, Grancharova T, Donovan-Maiye RM, Hendershott MC, Anderson HG, Brown JM, Chen J, Dinh SQ, Gehring JL, Johnson GR, et al. (2021). Cell states beyond transcriptomics: Integrating structural organization and gene expression in hiPSC-derived cardiomyocytes. *Cell Syst.* **12**, 670-687.e10.

Haas, P., and Gilmour, D. (2006). Chemokine Signaling Mediates Self-Organizing Tissue Migration in the Zebrafish Lateral Line. *Dev. Cell* **10**, 673–680.

Harding, M.J., and Nechiporuk, A.V. (2012). Fgfr-Ras-MAPK signaling is required for apical constriction via apical positioning of Rho-associated kinase during mechanosensory organ formation. *Development* **139**, 3130–3135.

Hartmann, J., Wong, M., Gallo, E., and Gilmour, D. (2020). An image-based data-driven analysis of cellular architecture in a developing tissue. *eLife* **9**, e55913.

Huff, J. (2015). The Airyscan detector from ZEISS: confocal imaging with improved signal-tonoise ratio and super-resolution. *Nat. Methods* **12**, i–ii.

Huttenlocher, D.P., Klanderman, G.A., and Rucklidge, W.J. (1993). Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863.

Itoh, M., and Chitnis, A.B. (2001). Expression of proneural and neurogenic genes in the zebrafish lateral line primordium correlates with selection of hair cell fate in neuromasts. *Mech. Dev.* **102**, 263–266.

Jiang, T., Kindt, K., and Wu, D.K. (2017). Transcription factor Emx2 controls stereociliary bundle orientation of sensory hair cells. *eLife*. **6**:e23661.

Kastan N, Gnedeva K, Alisch T, Petelski AA, Huggins DJ, Chiaravalli J, Aharanov A, Shakked A, Tzahor E, Nagiel A, et al. (2021). Small-molecule inhibition of Lats kinases may promote Yap-dependent proliferation in postmitotic mammalian tissues. *Nat. Commun.* **12**, 3100.

Keren, K., Pincus, Z., Allen, G.M., Barnhart, E.L., Marriott, G., Mogilner, A., and Theriot, J.A. (2008). Mechanism of shape determination in motile cells. *Nature* **453**, 475–480.

Kilian, K.A., Bugarija, B., Lahn, B.T., and Mrksich, M. (2010). Geometric cues for directing the differentiation of mesenchymal stem cells. *Proc. Natl. Acad. Sci.* **107**, 4872–4877.

Kimura, Y., Hisano, Y., Kawahara, A., and Higashijima, S. (2014). Efficient generation of knockin transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci. Rep.* **4**, srep06545.

Lambert, T., Liu, H., Buckley, G., Bokota, G., Stansby, D., Har-Gil, H., Nunez-Iglesias, J., Weigert, M., aeisenbarth, Selzer, G., Gay G, et al. (2023). pyapp-kit/magicgui: v0.7.2. Zenodo. 10.5281/zenodo.7700260.

Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el-AD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*. **162**:184-97.

López-Schier, H., Starr, C.J., Kappler, J.A., Kollmar, R., and Hudspeth, A.J. (2004). Directional Cell Migration Establishes the Axes of Planar Polarity in the Posterior Lateral-Line Organ of the Zebrafish. *Dev. Cell* **7**, 401–412.

Lush, M.E., Diaz, D.C., Koenecke, N., Baek, S., Boldt, H., St Peter, M.K., Gaitan-Escudero, T., Romero-Carvajal, A., Busch-Nentwich, E.M., Perera, A.G., et al. (2019). scRNA-Seq reveals distinct stem cell populations that drive hair cell regeneration after loss of Fgf and Notch signaling. *eLife* **8**, e44431.

Marklein, R.A., Lo Surdo, J.L., Bellayr, I.H., Godil, S.A., Puri, R.K., and Bauer, S.R. (2016). High Content Imaging of Early Morphological Signatures Predicts Long Term Mineralization Capacity of Human Mesenchymal Stem Cells upon Osteogenic Induction. *Stem Cells* **34**, 935–947.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861.

Millimaki, B.B., Sweet, E.M., Dhason, M.S., and Riley, B.B. (2007). Zebrafish atoh1 genes: classic proneural activity in the inner ear and regulation by Fgf and Notch. *Development* **134**, 295–305.

Pachitariu M, Stringer C. (2022). Cellpose 2.0: how to train your own model. *Nat Methods*. **19**:1634-1641.

Pasha Hosseinbor, A., Chung, M.K., Koay, C.G., Schaefer, S.M., van Reekum, C.M., Schmitz, L.P., Sutterer, M., Alexander, A.L., and Davidson, R.J. (2015). 4D hyperspherical harmonic (HyperSPHARM) representation of surface anatomy: A holistic treatment of multiple disconnected anatomical structures. *Med. Image Anal.* **22**, 89–101.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.

Perez-Vale KZ, Peifer M. (2020). Orchestrating morphogenesis: building the body plan by cell shape changes and movements. *Development*. **147**:dev191049.

Piccolo, S., Dupont, S., and Cordenonsi, M. (2014). The Biology of YAP/TAZ: Hippo Signaling and Beyond. *Physiol. Rev.* **94**: 1287-1312

Pincus, Z., and Theriot, J.A. (2007). Comparison of quantitative methods for cell-shape analysis. *J. Microsc.* **227**, 140–156.

Romero-Carvajal, A., Navajas Acedo, J., Jiang, L., Kozlovskaja-Gumbrienė, A., Alexander, R., Li, H., and Piotrowski, T. (2015). Regeneration of Sensory Hair Cells Requires Localized Interactions between the Notch and Wnt Pathways. *Dev. Cell* **34**, 267–282.

Rudolf, M.A., Andreeva, A., Kozlowski, M.M., Kim, C.E., Moskowitz, B.A., Anaya-Rocha, A., Kelley, M.W., and Corwin, J.T. (2020). YAP Mediates Hair Cell Regeneration in Balance Organs of Chickens, But LATS Kinases Suppress Its Activity in Mice. *J. Neurosci.* **40**, 3915– 3932.

Ruan, X., and Murphy, R.F. (2019). Evaluation of methods for generative modeling of cell and nuclear shape. *Bioinformatics* **35**, 2475–2485.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.

Smith, P.J., Wiltshire, M., Davies, S., Patterson, L.H., and Hoy, T. (1999). A novel cell permeant and far red-fluorescing DNA probe, DRAQ5, for blood cell discrimination by flow cytometry. *J. Immunol. Methods* **229**, 131–139.

Sofroniew, N., Lambert, T., Evans, K., Nunez-Iglesias, J., Bokota, G., Winston, P., PeñaCastellanos, G., Yamauchi, K., Bussonnier, M., Doncila Pop, D., et al. (2022). napari: a multi-dimensional image viewer for Python. doi:10.5281/zenodo.3555620

Thevenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41.

Thomas ED, Cruz IA, Hailey DW, Raible DW. (2015) There and back again: development and regeneration of the zebrafish lateral line system. *Wiley Interdiscip Rev Dev Biol.* **4**:1-16.

Thomas, E.D., and Raible, D.W. (2019). Distinct progenitor populations mediate regeneration in the zebrafish lateral line. *eLife* **8**, e43736.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing wellconnected communities. *Sci. Rep.* **9**, 5233.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in Python. *PeerJ* **2**, e453.

Viader-Llargués, O., Lupperger, V., Pola-Morell, L., Marr, C., and López-Schier, H. (2018). Live cell-lineage tracing and machine learning reveal patterns of organ regeneration. *eLife* **7**, e30823.

Viana, M.P., Chen, J., Knijnenburg, T.A., Vasan, R., Yan, C., Arakaki, J.E., Bailey, M., Berry, B., Borensztejn, A., Brown, E.M., et al. (2023). Integrated intracellular organization and its variations in human iPS cells. *Nature* **613**, 345–354.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* **17**, 261-272.

Watt, F.M., Jordan, P.W., and O'Neill, C.H. (1988). Cell shape controls terminal differentiation of human epidermal keratinocytes. *Proc. Natl. Acad. Sci.* **85**, 5576–5580.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15.

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59.

Yin Z, Sadok A, Sailem H, McCarthy A, Xia X, Li F, Garcia MA, Evans L, Barr AR, Perrimon N, Marshall CJ, Wong ST, Bakal C. (2013). A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat. Cell Biol.* **15**, 860– 871. 10.1038/ncb2764.

Figures

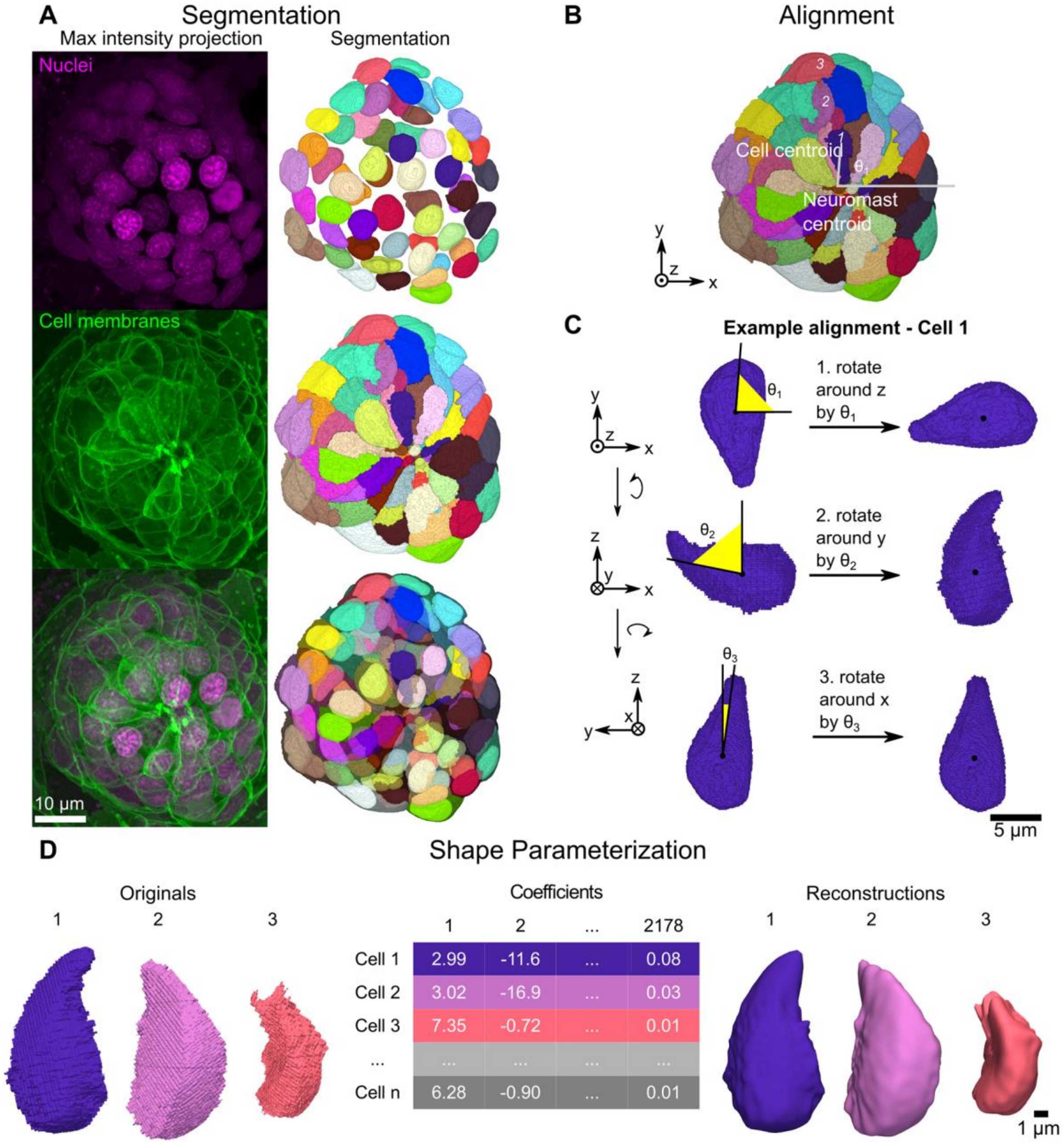
A
Segmentation
Max intensity projection
Segmentation
Nuclei
Cell membranes
10 µm
B
Alignment
3
2
1
Cell centroid
θ1
Neuromast centroid
y
z
x
C
Example alignment - Cell 1
θ1
1. rotate around z by θ1
y
z
x
θ2
2. rotate around y by θ2
z
y
x
θ3
3. rotate around x by θ3
z
x
y
5 µm
D
Shape Parameterization
Originals
1
2
3
Coefficients
1
2
...
2178
Cell 1
2.99
-11.6
...
0.08
Cell 2
3.02
-16.9
...
0.03
Cell 3
7.35
-0.72
...
0.01
...
Cell n
6.28
-0.90
...
0.01
Reconstructions
1
2
3
1 µm

Fig. 1. Overview of segmentation and shape parameterization

(A) Representative maximum intensity projections (left column) and corresponding 3D projections (right column) for nuclei labeled by DRAQ5 dye (top row), cells labeled by Tg(-8.0cldnb:LY-EGFP) (middle row), and both channels together (bottom row).

(B) Image depicting how the first rotation angle is calculated. Example given for the maximum intensity projection segmentation channel of the neuromast shown in (A). Cells 1, 2, and 3 are used in (C) and (D).

(C) Example of the alignment procedure as applied to cell 1 from (B). To correct for radial symmetry, cells are first rotated around the z axis. Cells were then rotated around the y axis to correct for apicobasal tilt, and then around the x axis to correct for off-parallel tilt during imaging.

(D) Overview of shape parameterization using SH, as applied to cells 1, 2, and 3 from (B). SH expansion is applied to original 3D surface meshes (left) to generate the cell-by-SHE coefficients table (middle) which can then be used to generate 3D reconstructions (right).

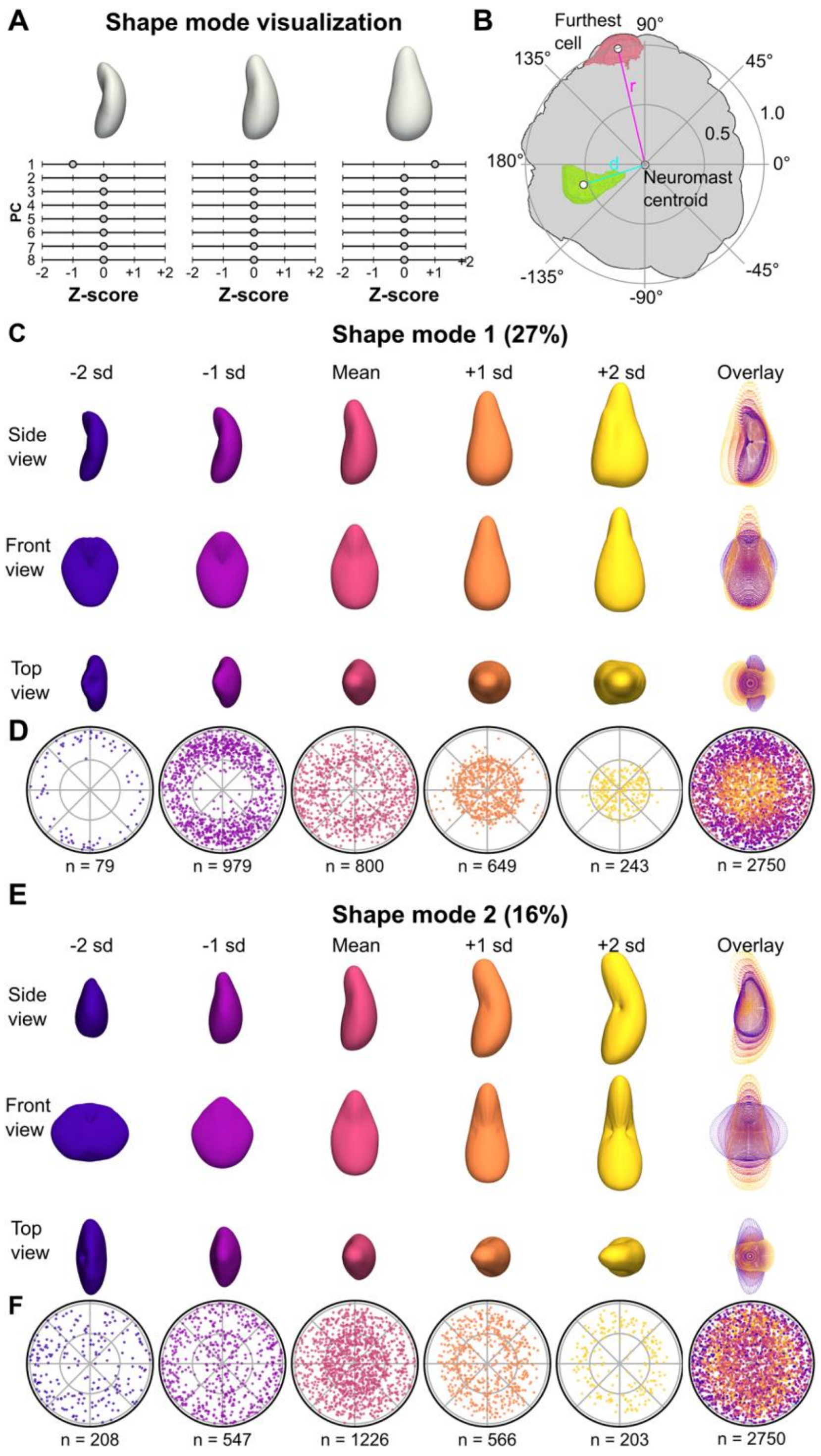

Fig. 2. Visualization of cell shape modes 1 and 2

(A) Schematic depicting how shape variation was visualized. PCs were first z-scored to generate shape modes. Visual representations of each shape mode were generated by finding the mean cell (a hypothetical cell with a value of 0 for all 8 shape modes) and varying one shape mode (here, shape mode 1) by up to 2 standard deviations (sd) while holding other shape modes at the mean (0).

(B) Schematic depicting how relative cell locations within each neuromast were calculated. The distance between the neuromast centroid and the furthest cell (r) in the XY plane was calculated. The distances between the neuromast centroid and cell centroids were divided by (r) to yield the normalized distance from the neuromast center (d). The angular location of each cell was determined in reference to the positive x-axis centered on the neuromast centroid. Cell locations were projected onto a unit circle with radius 1.

(C) Visualization of CSM1, generated as described in (A). 3D surface renderings are shown from 3 different views, as well as an overlay of point representations for each view (rightmost column).

(D) Polar plots of relative cell locations (generated as shown in B) binned by CSM1 scores. Bins were centered on the corresponding z-score and included cells with values within 0.5. For the -2 sd and 2 sd plots, the upper bounds were not limited to within 0.5 (e.g. the first plot includes all cells with CSM1 values < -1.5 sd). An overlay of the first 5 plots is shown in the rightmost column.

(E) Visualization of CSM2. 3D surface renderings are shown from 3 different views, as well as an overlay of point representations for each view (rightmost column).

(F) Polar plots of relative cell locations binned by CSM2 scores. An overlay of the first 5 plots is shown in the rightmost column.

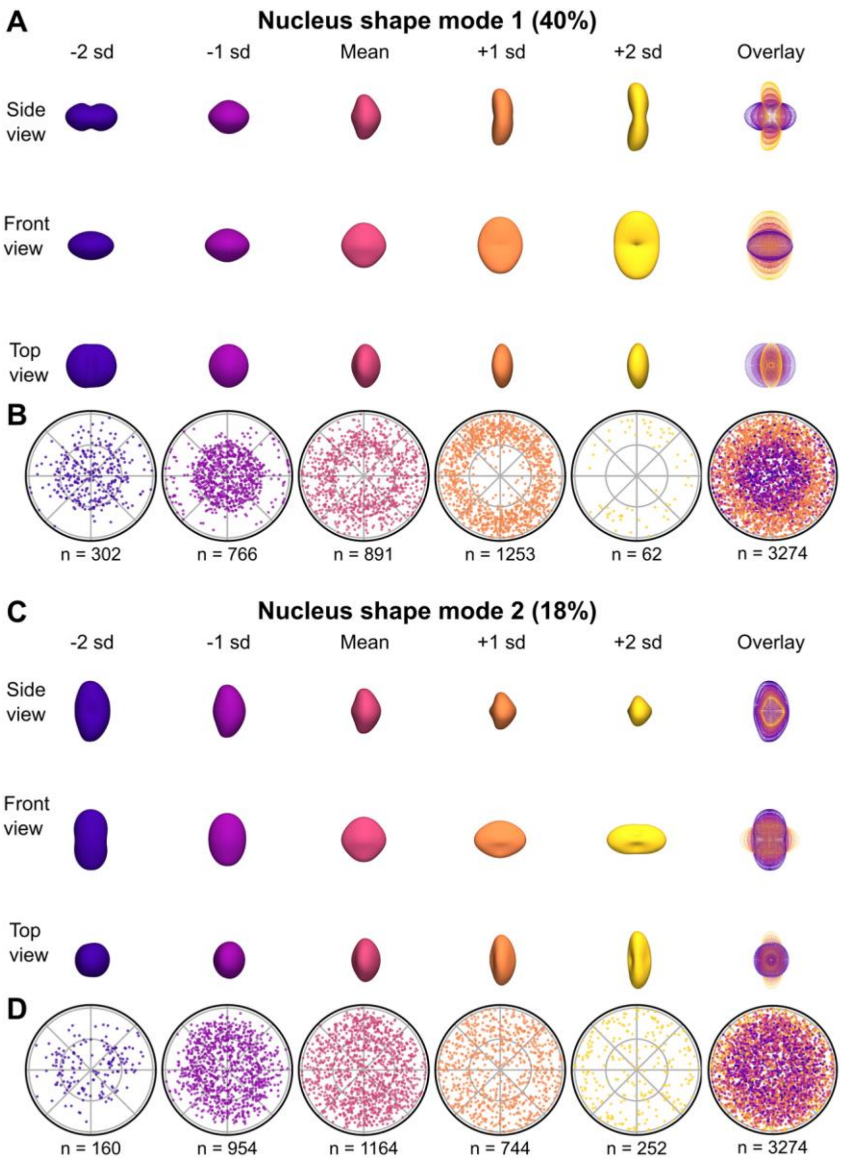

Fig. 3. Visualization of nuclear shape modes 1 and 2

(A) Visualization of NSM1, generated as in Figure 2A. 3D surface renderings are shown from 3 different views, as well as an overlay of point representations for each view (rightmost column).

(B) Polar plots of relative cell locations (generated as in Figure 2D) binned by NSM1 scores. An overlay of the first 5 plots is shown in the rightmost column.

(C) Visualization of NSM2. 3D surface renderings are shown from 3 different views, as well as an overlay of point representations for each view (rightmost column).

(D) Polar plots of relative cell locations binned by NSM2 scores. An overlay of the first 5 plots is shown in the rightmost column.

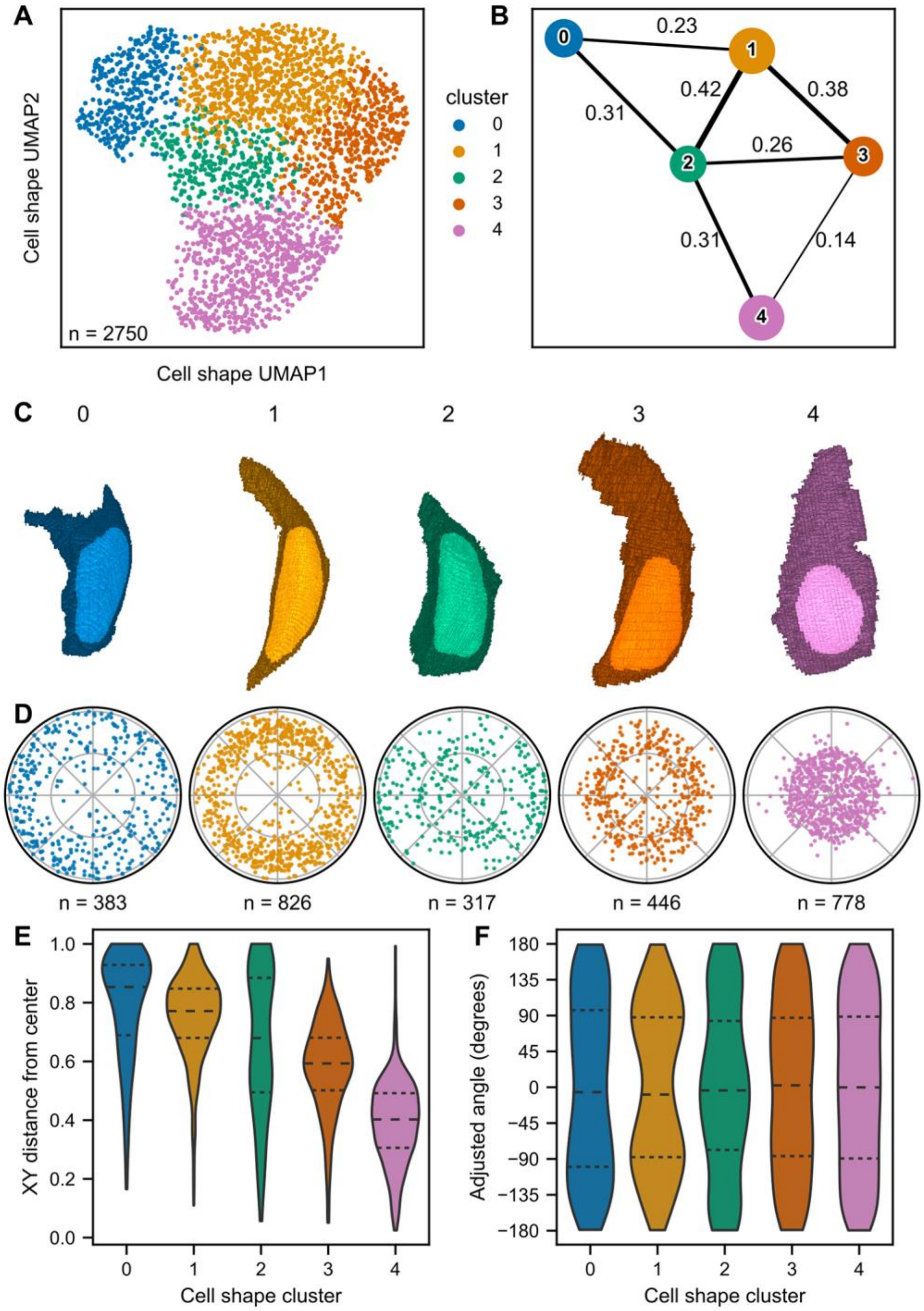

Fig. 4. Unsupervised clustering of cells in shape space

(A) UMAP of neuromast cells, color coded by cell shape cluster. Colors and associated cluster numbers are used in (B-F).

(B) PAGA plot of cell shape clusters. Each node represents a cluster. Thickness of the line drawn between nodes indicates estimated connectivity between clusters (the value of which is shown adjacent to the corresponding line). Edges with connectivity below 0.1 are not depicted.

(C) 3D projections of representative cells for each cluster, defined as the cell closest to the cluster centroid in eight-dimensional PC space.

(D) Polar plots of relative neuromast locations of cells within each cluster.

(E) Distributions of the cell distance from the neuromast center for each cluster. Dashed lines indicate quartiles.

(F) Distributions of neuromast cell angles for each cluster. Dashed lines indicate quartiles. Angles were calculated with reference to the positive x-axis extending from the neuromast centroid. Cluster 1 shows a bias towards the dorsoventral poles (90, -90), while cluster 2 shows a bias to the anterior-posterior compartments (0, 180, -180).

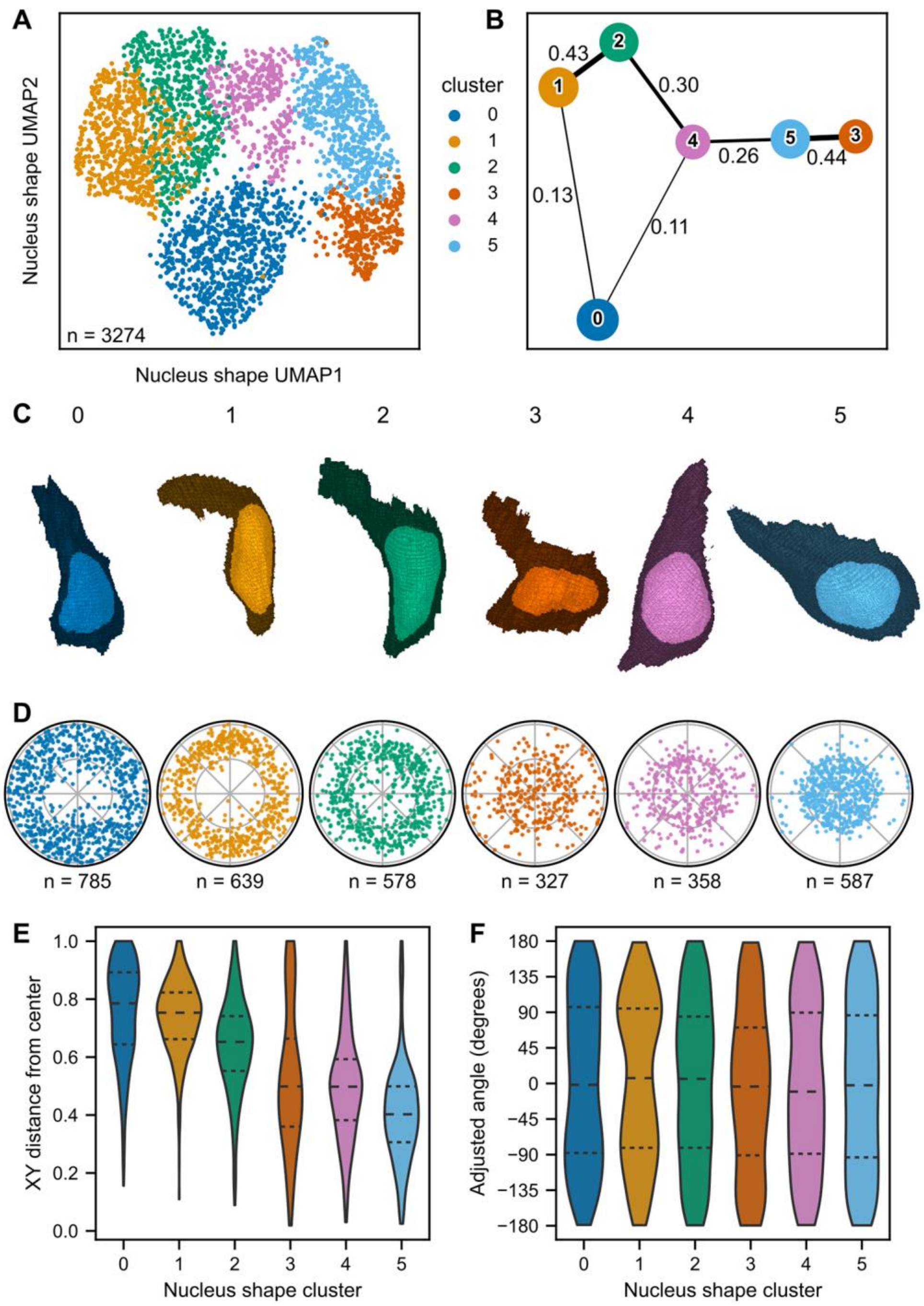

Fig. 5. Unsupervised clustering of nuclei in shape space

(A) UMAP of neuromast cells, color coded by nucleus shape cluster. Colors and associated cluster numbers are used in (B), (C), (D), (E), and (F).

(B) PAGA plot of nucleus shape clusters. Each node represents a cluster. Thickness of the line drawn between nodes indicates estimated connectivity between clusters (the value of which is shown adjacent to the corresponding line). Edges with connectivity below 0.1 are not depicted.

(C) 3D projections of representative cells for each nuclear cluster, defined as the cell closest to the cluster centroid in four-dimensional PC space. The original segmentation masks are depicted (not SHE reconstructions).

(D) Polar plots of relative locations of cells within each cluster.

(E) Distributions of the cell distance from the neuromast center for each cluster. Dashed lines indicate quartiles.

(F) Distributions of neuromast cell angles for each cluster. Dashed lines indicate quartiles. Angles were calculated with reference to the positive x-axis extending from the neuromast centroid.

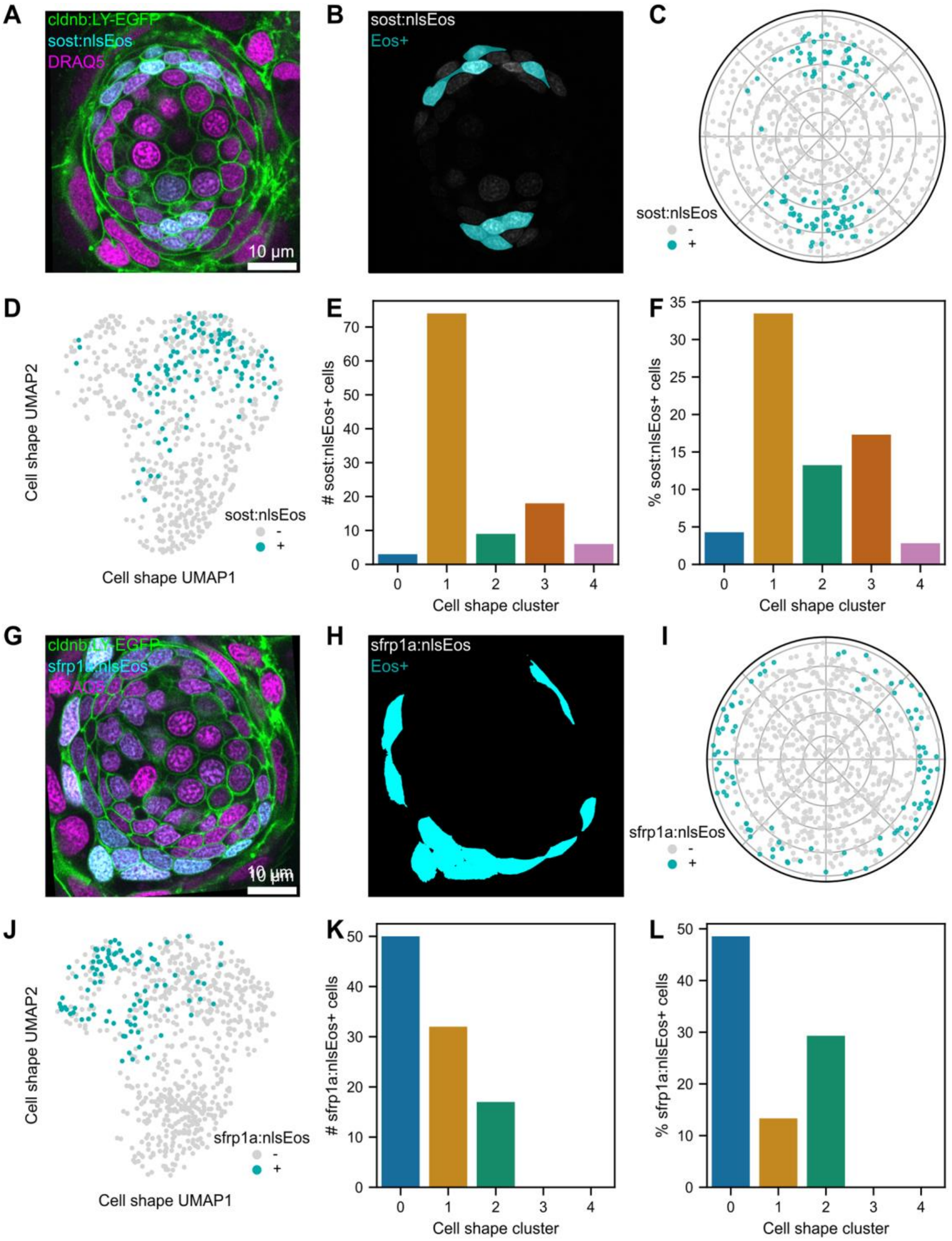

Fig. 6. Location and shape characteristics of cells expressing *sost:NLS-Eos* and sfrp1a:NLS-Eos

(A) Single confocal slice of a neuromast from sost:NLS-Eos; Tg(-8.0cldnb:LY-EGFP) fish stained with DRAQ5. Image contrast was adjusted for visibility.

(B) The same slice as in (A), with the sost:NLS-Eos channel (gray) overlaid with a mask indicating cells classified as Eos+ (dark cyan).

(C) Polar plot showing relative cell locations of locations of Eos+ (dark cyan) and Eos- (gray) cells.

(D) UMAP of Eos+ (dark cyan) and Eos- (gray) cells from sost:NLS-Eos.

(E) sost:NLS-Eos+ cells plotted as raw counts within each cell shape cluster.

(F) sost:NLS-Eos+ cells plotted as the percentage of each cluster.

(G) Single confocal slice of a neuromast from sfrp1a:NLS-Eos; Tg(-8.0cldnb:LY-EGFP) fish stained with DRAQ5. Image contrast was adjusted for visibility.

(H) The same slice as in (A), with the sfrp1a:NLS-Eos channel (gray) overlaid with a mask indicating cells classified as Eos+ (dark cyan).

(I) Polar plot showing relative cell locations of locations of Eos+ (dark cyan) and Eos- (gray) cells.

(J) UMAP of Eos+ (dark cyan) and Eos- (gray) cells from sfrp1a:NLS-Eos.

(K) sfrp1a:NLS-Eos+ cells plotted as raw counts within each cell shape cluster.

(L) sfrp1a:NLS-Eos+ cells plotted as the percentage of each cluster.

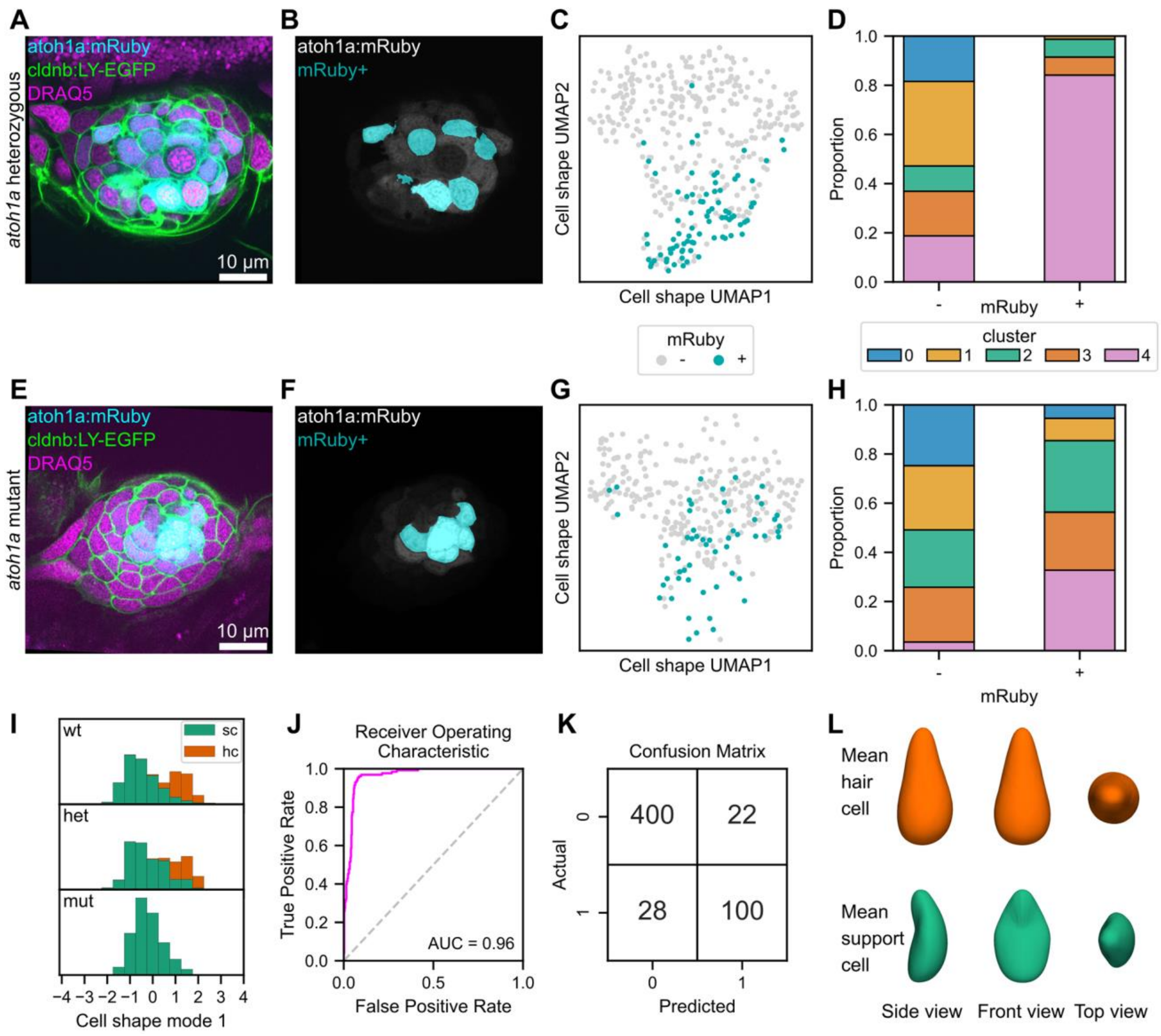

Fig. 7. Shape characteristics of *atoh1a:mRuby* heterozygotes and mutants and logistic classifier for hair cells based on cell shape features

Analysis of cell shape distributions in atoh1a:mRuby heterozygotes (A-D) and mutants (E-G).

(A, E) Single confocal slice of a neuromast from a atoh1a:mRuby heterozygote (A) and mutant (E) also expressing Tg(-8.0cldnb:LY-EGFP) and stained with DRAQ5. Image contrast was adjusted for visibility.

(B, F) The same slices as in (A) and (E), with the atoh1a:mRuby channel (gray) overlaid with a mask indicating cells classified as mRuby+ (dark cyan) for a atoh1a:mRuby heterozygote (B) and mutant (F).

(C, G) UMAP of mRuby+ (dark cyan) and mRuby- (gray) cells from atoh1a:mRuby heterozygotes (C) and mutants (F).

(D, H) Proportions of mRuby- cells (left bar) and mRuby+ cells (right bar) from atoh1a:mRuby heterozygotes (D) and mutants (H) in each cell shape cluster.

(I) Distributions of cell shape mode 1 scores, separated by genotype (top row: wildtype; middle row: atoh1a heterozygous; bottom row: atoh1a mutant). Hair cells (hc, green); support cells (sc, orange).

(J) Receiver Operating Characteristic (ROC) curve (fuchsia) for a logistic regression classifier trained to detect hair cells based on cell shape mode 1-8. The Area Under the Curve (AUC) score is shown (0.96). Silver dashed line = performance of a random classifier.

(K) Confusion matrix for the cell shape-based hair cell classifier. 0 = support cells; 1 = hair cells. Top left: true negatives; top right: false positives; bottom left: false negatives; bottom right: true positives.

(L) The idealized mean cell shape for hair cells (top, orange) and support cells (bottom, green).

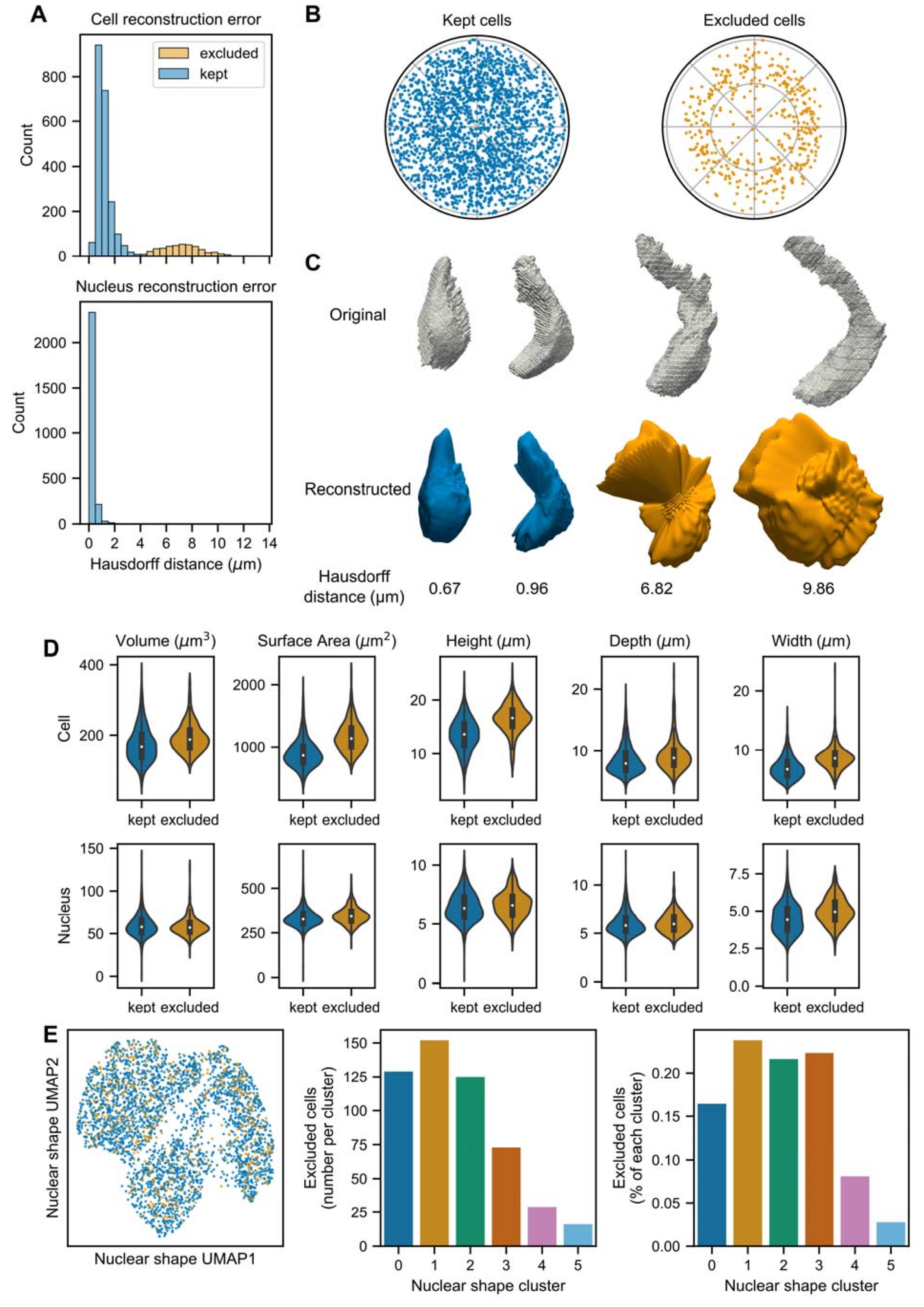
A
Cell reconstruction error
excluded
kept
Count
0
200
400
600
800
Nucleus reconstruction error
Count
0
500
1000
1500
2000
0 2 4 6 8 10 12 14
Hausdorff distance (μm)
B
Kept cells
Excluded cells
C
Original
Reconstructed
Hausdorff distance (μm)
0.67
0.96
6.82
9.86
D
Volume (μm³)
Surface Area (μm²)
Height (μm)
Depth (μm)
Width (μm)
Cell
Nucleus
kept excluded
E
Nuclear shape UMAP1
Nuclear shape UMAP2
Excluded cells (number per cluster)
Excluded cells (% of each cluster)
Nuclear shape cluster
0 1 2 3 4 5

Fig. S1.

(A) Distribution of reconstruction error, as measured by directed Hausdorff distance, for cell shape (top) and nuclear shape (bottom). Some cells were excluded from cell shape analysis (orange) due to high reconstruction error.

(B) Polar plots showing relative locations of cells kept (left, blue) and excluded from cell shape analysis (right, orange).

(C) Examples of well reconstructed (left) and poorly reconstructed (right) cells. 3D representations of the original mesh (top), reconstructed mesh (middle), and the associated Hausdorff distances (bottom) for two examples of each are shown.

(D) Distributions of cell (top) and nuclei (bottom) shape features in kept (blue) and excluded (orange) cells.

(E) Left: UMAP projection of cells in nuclear shape space (corresponding with Figure 5A) colored by whether they were kept (blue) or excluded (orange) from cell shape analysis. Middle: Excluded cells plotted as number

(raw counts) for each nuclear shape cluster. Right: Excluded cells as a percentage of each nuclear shape cluster.

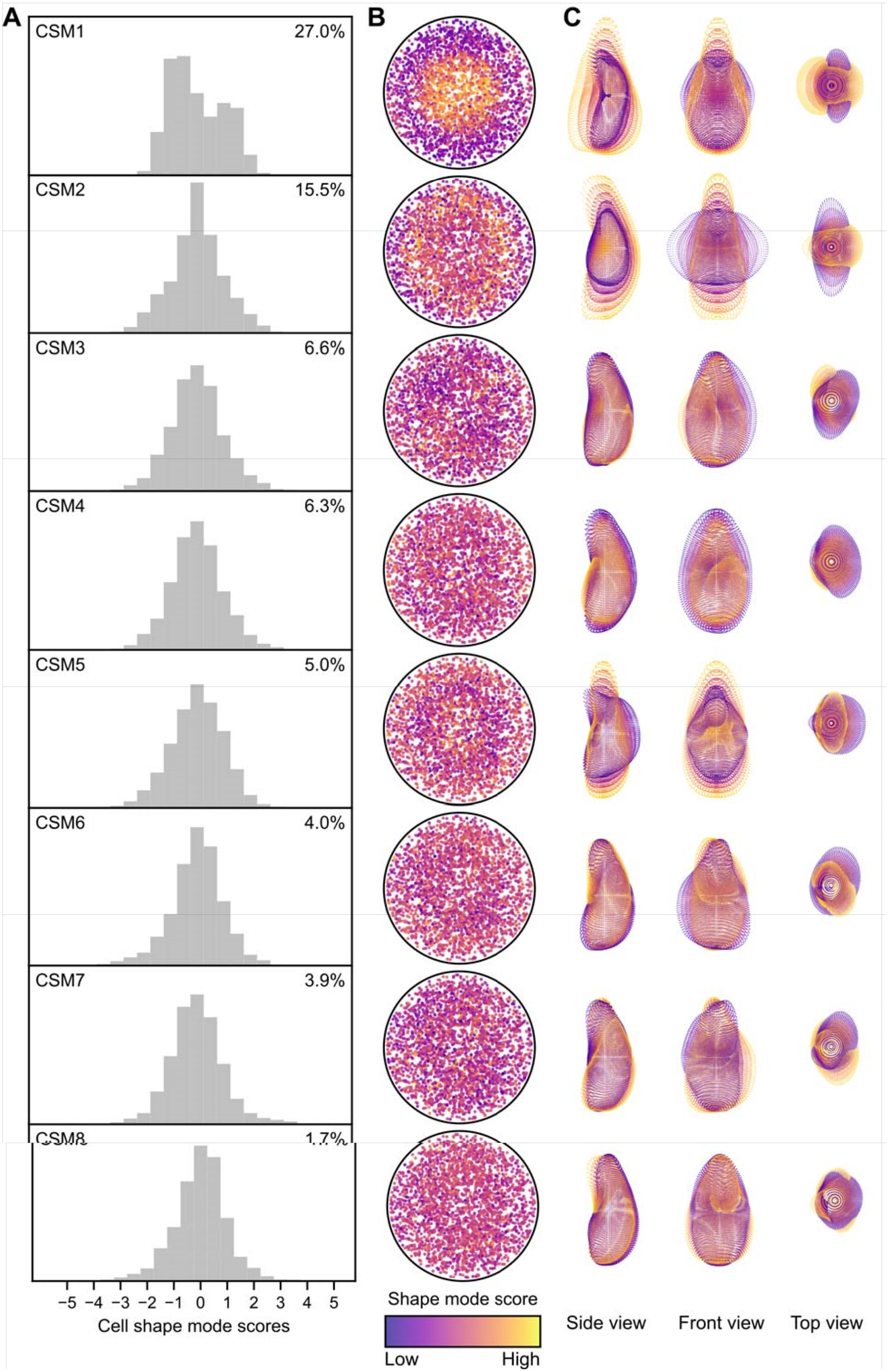
A
CSM1
27.0%
CSM2
15.5%
CSM3
6.6%
CSM4
6.3%
CSM5
5.0%
CSM6
4.0%
CSM7
3.9%
CSM8
1.7%
−5
−4
−3
−2
−1
0
1
2
3
4
5
Cell shape mode scores
B
Shape mode score
Low
High
C
Side view
Front view
Top view

Fig. S2.

(A) Distributions of shape mode scores for the first eight cell shape modes (CSM1-CSM8). Percent variance explained for each shape mode is provided in the upper right corner.

(B) Polar plots showing relative locations of neuromast cells in the dataset, color coded by shape mode scores for CSM1-CSM8.

(C) Overlaid 3D point view representations of CSM1-CSM8, generated by finding the mean cell and varying each shape mode by up to 2 standard deviations (sd) in increments of 1 sd while holding other shape modes at the mean (0).

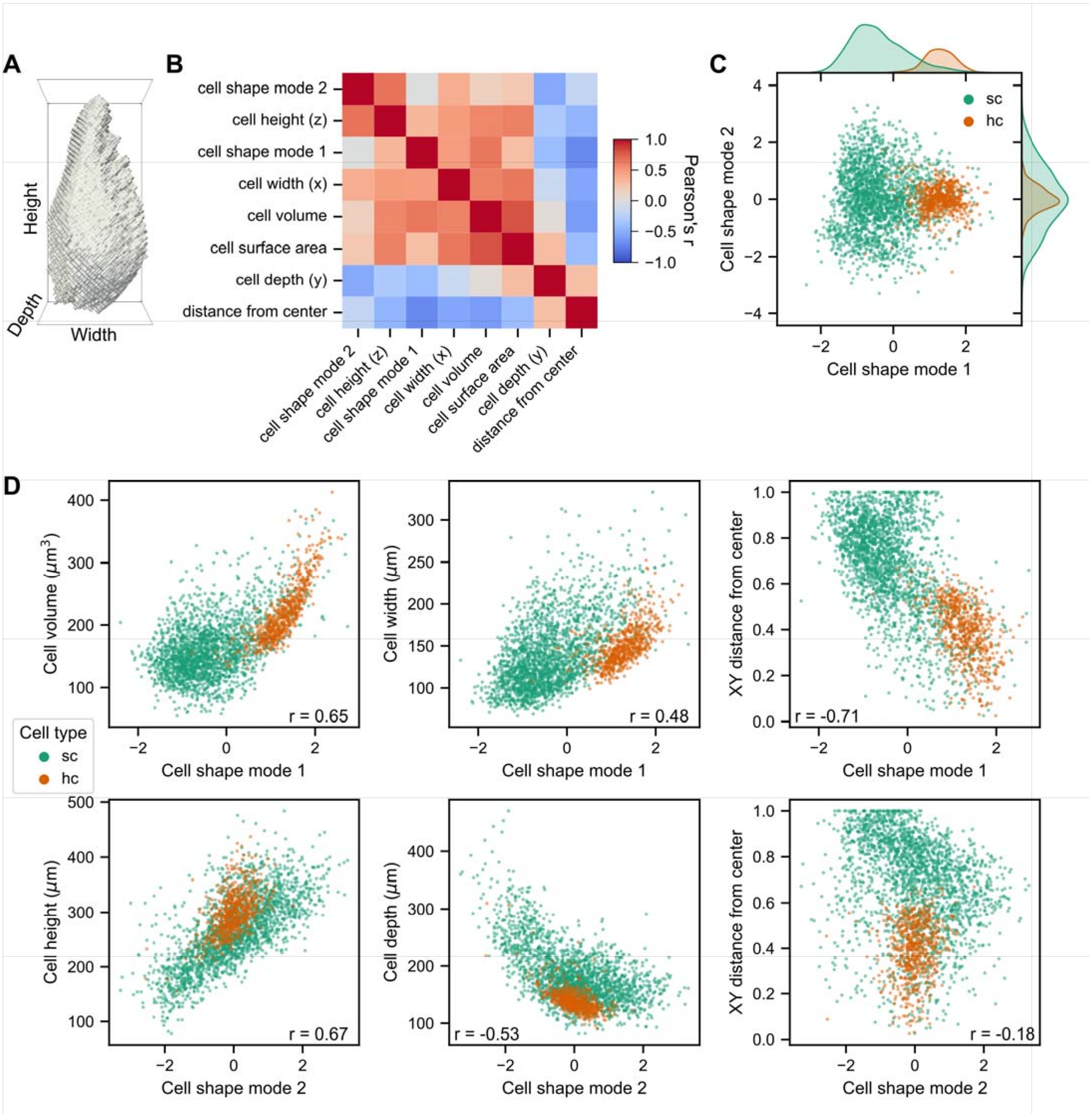
A
Height
Depth
Width
B
cell shape mode 2
cell height (z)
cell shape mode 1
cell width (x)
cell volume
cell surface area
cell depth (y)
distance from center
1.0
0.5
0.0
−0.5
−1.0
Pearson's r
C
Cell shape mode 2
Cell shape mode 1
sc
hc
D
Cell volume (μm³)
Cell width (μm)
Cell height (μm)
Cell depth (μm)
XY distance from center
r = 0.65
r = 0.48
r = -0.71
r = 0.67
r = -0.53
r = -0.18
Cell type
sc
hc

Fig. S3.

(A) Side view of a 3D projection of a representative cell with its dimensions labeled.

(B) Heatmap of correlations between cell shape mode 1, cell shape mode 2, and cell shape/location features.

(C) Plot of cell shape mode 1 and cell shape mode 2 for SCs (green) and hair cells (orange).

(D) Top row: Correlations between cell volume (left), width (center), and distance from center (right) versus cell shape mode 1 color coded by cell type. Bottom row: Correlations between cell height (left), depth (center), and distance from center (right) versus cell shape mode 2 color coded by cell type. The Pearson correlation coefficient (r) is displayed on each plot. SCs (green) and hair cells (orange).

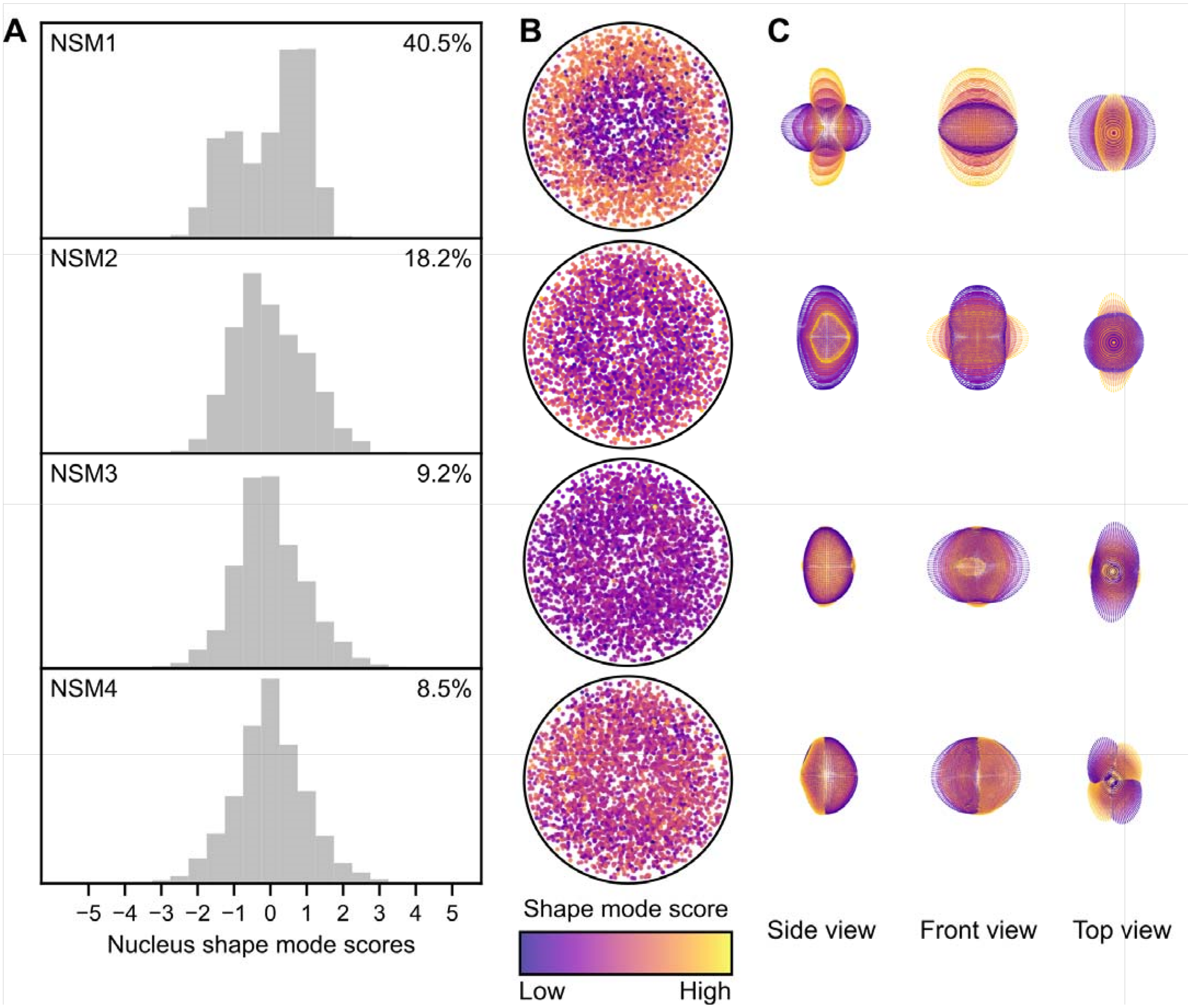

Fig. S4.

(A) Distributions of shape mode scores for the first four nucleus shape modes (NSM1-NSM4). Percent variance explained for each shape mode is provided in the upper right corner.

(B) Polar plots showing relative locations of neuromast cells in the dataset, color coded by shape mode scores for NSM1-NSM4.

(C) Overlaid 3D point view representations of CSM1-CSM8, generated by finding the mean cell and varying each shape mode by up to 2 standard deviations (sd) in increments of 1 sd while holding other shape modes at the mean (0).

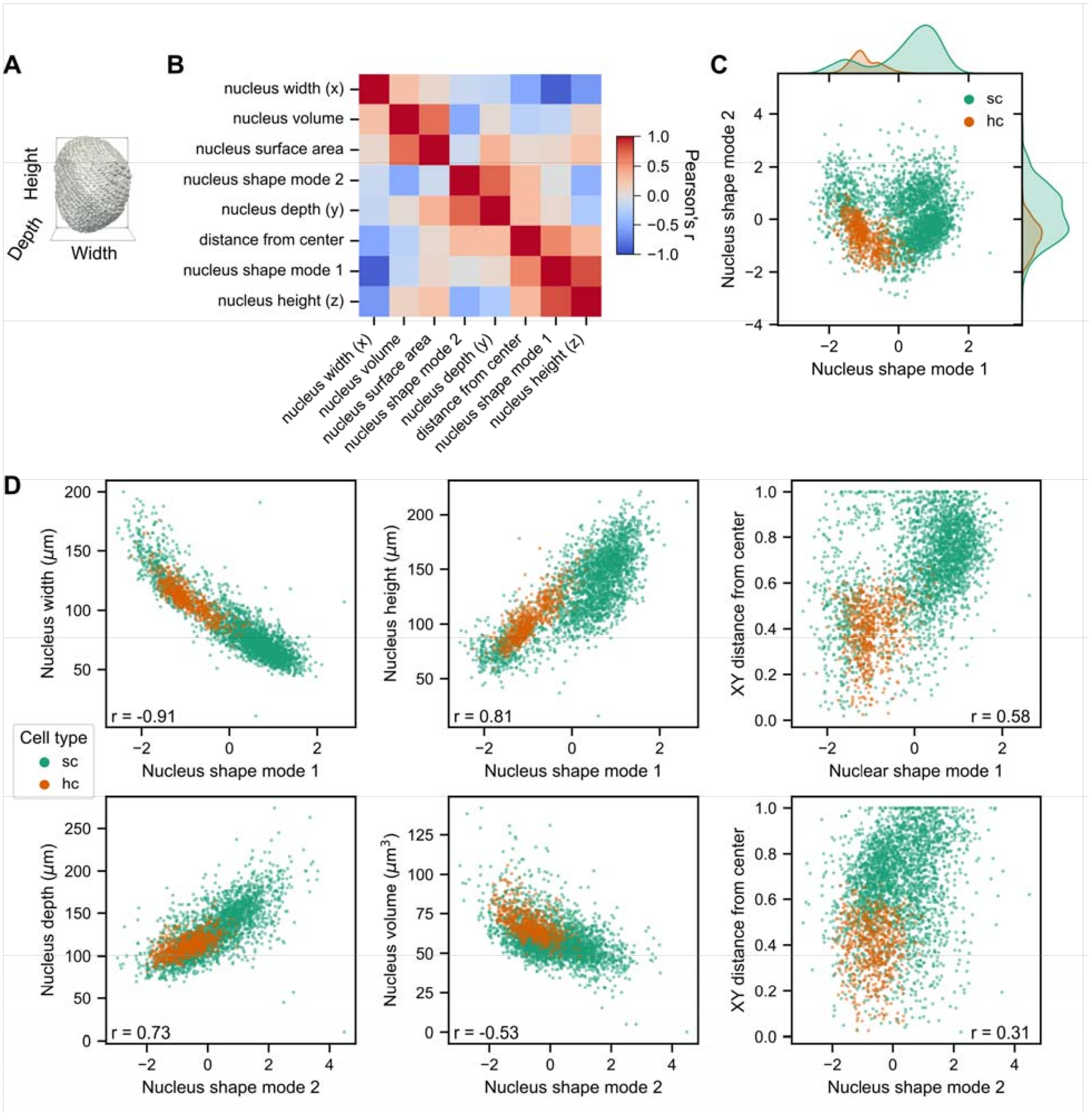

Fig. S5.

(A) Side view of a 3D projection of a representative nucleus with its dimensions labeled.

(B) Heatmap of correlations between nucleus shape mode 1, nucleus shape mode 2, and nucleus shape/location features.

(C) Plot of nucleus shape mode 1 and nucleus shape mode 2 for SCs (green) and hair cells (orange).

(D) Top row: Correlations between nucleus width (left), height (center), and distance from center (right) versus nucleus shape mode 1 color coded by cell type. Bottom row: Correlations between nucleus depth (left), volume (center), and distance from center (right) versus nucleus shape mode 2 color coded by cell type. The Pearson correlation coefficient (r) is displayed on each plot. SCs (green) and hair cells (orange).

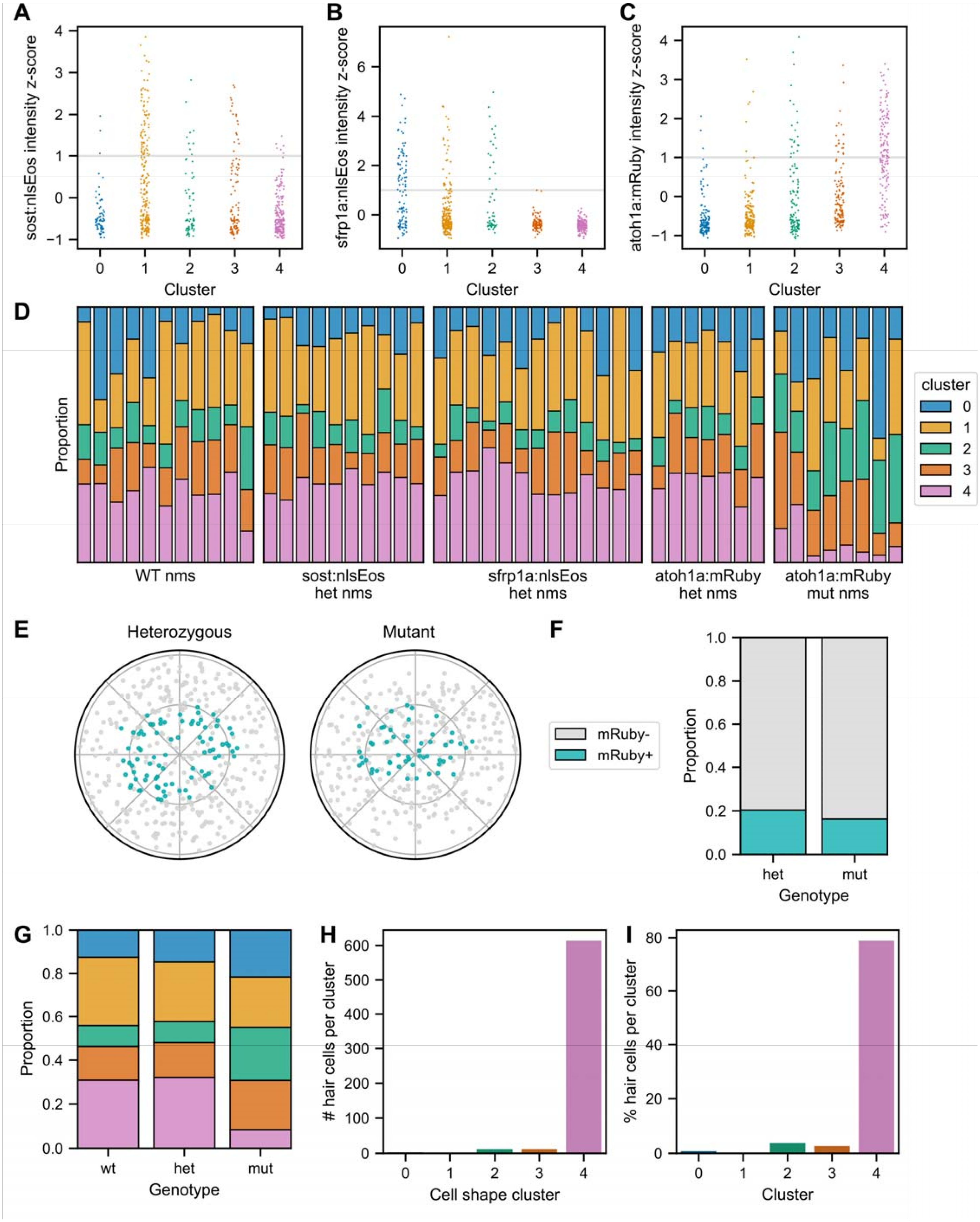
A
sost:nlsEos intensity z-score
Cluster
0
1
2
3
4
B
sfrp1a:nlsEos intensity z-score
C
atoh1a:mRuby intensity z-score
D
Proportion
WT nms
sost:nlsEos het nms
sfrp1a:nlsEos het nms
atoh1a:mRuby het nms
atoh1a:mRuby mut nms
cluster
E
Heterozygous
Mutant
mRuby-
mRuby+
F
Proportion
het
mut
Genotype
G
Proportion
wt
het
mut
Genotype
H
# hair cells per cluster
Cell shape cluster
I
% hair cells per cluster
Cluster

Fig. S6.

(A) Distribution of mean *sost:NLS-Eos* intensities (z-scored) for cells in each cell shape cluster. Gray line represents intensity threshold (1 SD) used to label a cell as positive.

(B) Distribution of mean *sfrp1a:NLS-Eos* intensities (z-scored) for cells in each cell shape cluster. Gray line represents intensity threshold (1 SD) used to label a cell as positive.

(C) Distribution of mean *atoh1a:mRuby* intensities (z-scored) for cells in each cell shape cluster. Gray line represents intensity threshold (1 SD) used to label a cell as positive.

(D) Proportions of cells in individual neuromasts belonging to each cluster, grouped by reporter expression and genotype (n=49 neuromasts).

(E) Polar plots of relative cell locations of mRuby- cells (gray) and mRuby+ cells (dark cyan) in *atoh1a:mRuby* heterozygotes (left) and *atoh1a:mRuby* mutants (right).

(F) Proportions of mRuby+ cells (dark cyan) and mRuby- cells (gray) in *atoh1a:mRuby* heterozygotes (left bar) and *atoh1a:mRuby* mutants (right bar).

(G) Proportions of cells of each genotype - *atoh1a:mRuby* mutants, heterozygotes, and wildtype - belonging to each cell shape cluster.

(H) Hair cells as raw counts (number) within each cell shape cluster.

(I) Hair cells as a percentage of each cell shape cluster.

Table S1. Summary of the dataset. Each batch represents a separate imaging session. All neuromasts were labeled with DRAQ5 dye in the *Tg(-8.0cldnb:LY-EGFP)* background; the "reporter" column indicates additional genetic labels. "Filter" refers to cells excluded for having high cell shape reconstruction error (see Fig. S1 for details).

batch	reporter	# neuromasts	# cells (pre-filter)	# cells (post-filter)
1	sfrp1a:nlsEos	7	453	383
2	sost:nlsEos	10	824	676
3	sfrp1a:nlsEos	6	459	371
4	atoh1a:mRuby	10	560	496
5	atoh1a:mRuby	5	295	244
6	N/A (wildtype)	11	683	580