

## DEVELOPMENTAL BIOLOGY

## Crustacean leg regeneration restores complex microanatomy and cell diversity

Alba Almazán<sup>1†</sup>, Çağrı Çevrim<sup>1†</sup>, Jacob M. Musser<sup>2</sup>, Michalis Averof<sup>1\*</sup>, Mathilde Paris<sup>1\*</sup>

Animals can regenerate complex organs, yet this process frequently results in imprecise replicas of the original structure. In the crustacean *Parhyale*, embryonic and regenerating legs differ in gene expression dynamics but produce apparently similar mature structures. We examine the fidelity of *Parhyale* leg regeneration using complementary approaches to investigate microanatomy, sensory function, cellular composition, and cell molecular profiles. We find that regeneration precisely replicates the complex microanatomy and spatial distribution of external sensory organs and restores their sensory function. Single-nuclei sequencing shows that regenerated and uninjured legs are indistinguishable in terms of cell-type composition and transcriptional profiles. This remarkable fidelity highlights the ability of organisms to achieve identical outcomes via distinct processes.

## INTRODUCTION

The ability to regenerate varies widely among animals. On one extreme of the spectrum, planarians and hydrozoans are thought to be capable of perfect regeneration, coupling fission with regeneration to achieve asexual reproduction (1). On the other extreme are organisms in which regeneration is lacking (e.g., nematodes) or limited to physiological cell turnover in tissues such as epithelia and blood. Between these two extremes are a wide variety of animals with imperfect regeneration. Examples include *Xenopus* froglets that replace amputated limbs by poorly patterned spikes (2), lizards that replace the bony vertebrae of their tails with cartilage (3, 4), and cockroach nymphs that regenerate legs that have small but consistent differences in nerves, musculature, and tracheae compared with unharmed legs (5). In some cases, including well-studied axolotl legs and zebrafish fins, regeneration produces organs of normal appearance that may carry subtle defects (6–9). In most instances of regeneration, it is unknown whether the cellular composition, detailed morphology, and function of regenerated organs are fully restored.

We approached this question in the crustacean *Parhyale hawaiiensis*, an experimental system in which legs typically regenerate within 1 to 2 weeks after amputation (10, 11). Regenerated *Parhyale* legs are smaller than control legs in the first molt following amputation but recover in size gradually during subsequent molts (11). On casual inspection, regenerated legs appear identical to uninjured ones in terms of leg segment morphology and the presence of muscles and motor neurons (10). Despite this, *Parhyale* leg regeneration and leg development differ in their global gene expression dynamics (12), suggesting that the processes involved in generating legs in the embryo and regenerating legs in the adult stage are likely to differ. We used microanatomical observations, a mechanosensory assay, and single-nuclei transcriptomics to address whether regenerated legs differ from their unharmed counterparts in terms of morphology, sensory function, and cell-type diversity and molecular profiles.

## RESULTS

## Diversity and distribution of external sensory organs

To survey fidelity in regenerated legs, we first focused on the recovery of different types of external sensory organs, exposed as setae on the surface of the legs. These structures provide exquisite markers for assessing the accuracy of regeneration with high—almost cellular—precision, because their detailed morphologies and spatial patterns are complex and reproducible among individuals. Similar to insect sensory bristles, setae are composed of a few cells and take different shapes depending on their type and function (13, 14). Their complex and stereotypic external morphology enables easy detection of even minor defects.

Scanning electron microscopy (SEM) on the three most distal segments of uninjured *Parhyale* legs revealed eight types of setae with distinct morphologies (Fig. 1, fig. S1, and table S1). Immunostainings of acetylated  $\alpha$ -tubulin, which label nerve axons, showed that all the setae are innervated, as expected of sensory organs (Fig. 1B and fig. S2). Each setal type has a specific distribution on T4 and T5 thoracic legs, either as a single seta in a stereotypic location (e.g., single hooked, curved, and plumose setae in the dactylus and single cuspidate seta in the propodus) or in groups of setae arranged in specific spatial patterns (e.g., crowns, combs, and ventral arrays made of lamellate and twin setae in the propodus and carpus) (Fig. 1, A and C, and figs. S3 and S4). In these latter types, the number of setae varies depending on podomere size (see below, fig. S5A). The only exception to these reproducible spatial arrangements is seen with the microsetae, which are spaced on the surface of the cuticle without an apparent conserved pattern (Fig. 1A).

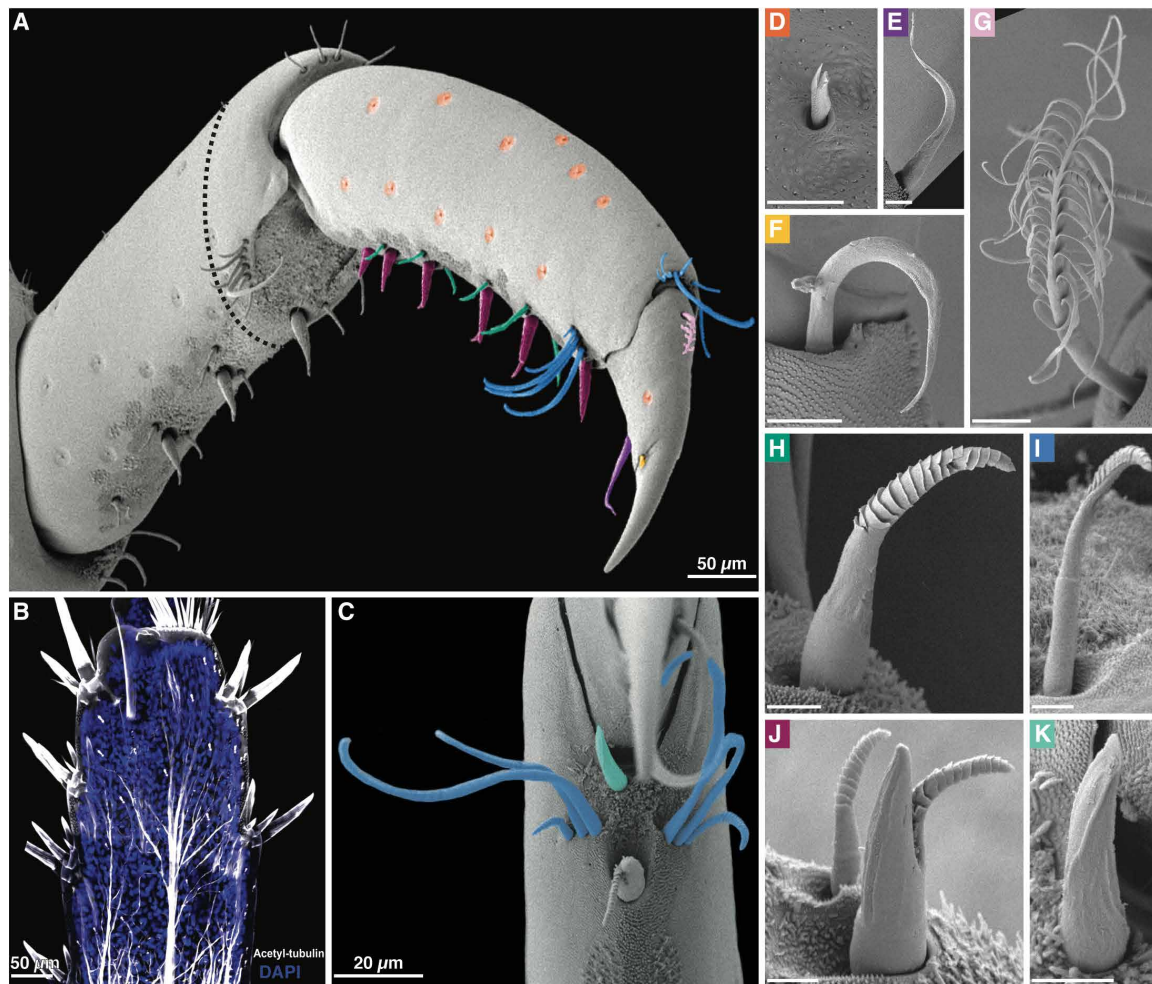
We then examined whether the diversity, morphology, and spatial distribution of these sensory organs are restored during regeneration. To minimize genetic, environmental, and physiological effects, we compared the setal patterns between regenerated and control (uninjured contralateral) legs from the same individuals. We found that the detailed morphologies and the spatial distributions of all setal types are unchanged in regenerated legs (fig. S1, table S1, and data file S1). In particular, the unique setae found in stereotypic positions (hooked, curved, plumose, and cuspidate setae) are consistently regenerated in their original locations (fig. S1 and data file S1). Setal types with variable numbers (lamellate type 1 and type 2 and twin setae in the crown, combs, and ventral arrays) are less abundant at the first molt after regeneration but recover

Copyright © 2022  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Institut de Génomique Fonctionnelle de Lyon (IGFL), Centre National de la Recherche Scientifique (CNRS), École Normale Supérieure de Lyon and Université Claude Bernard Lyon 1, 69007 Lyon, France. <sup>2</sup>European Molecular Biology Laboratory, Developmental Biology Unit, Meyerhofstrasse 1, Heidelberg 69117, Germany.

\*Corresponding author. Email: michalis.averof@ens-lyon.fr (M.A.); mathilde.paris@ens-lyon.fr (M.P.)

†These authors contributed equally to this work.



**Fig. 1. Diversity of external sensory organs in *Parhyale* legs.** (A) Distal part of an intact *Parhyale* T5 leg, viewed posteriorly by SEM. Different types of external sensory organs (setae) in the propodus and dactylus are highlighted in different colors [details shown in (D) to (K)]. The amputation plane in the setal regeneration experiments is indicated by a black dashed line. (B) Acetylated  $\alpha$ -tubulin staining of the carpus of a T6 leg showing the innervation of the setae. DAPI, 4',6'-diamidino-2-phenylindole. (C) Ventral view of the distal part of the propodus and the dactylus (out of focus) of T5 leg, by SEM, showing the anterior and posterior combs made of lamellate type 2 setae (cyan) and a single cuspidate seta (turquoise). (D to K) High-magnification view of each type of seta found on the distal part of *Parhyale* T4 and T5 legs: microseta (D), curved seta (E), plumose seta (G), lamellate type 1 seta (H), lamellate type 2 seta (I), and cuspidate seta (K). Panel labels match the coloring of setae in (A) and (C). Additional descriptions are given in the Materials and Methods and figs. S1 and S3. Scale bars, 5  $\mu$ m (D to K).

fully during subsequent molts [Fig. 2B (right) and fig. S5B]. This mirrors leg size, which is initially smaller in regenerated legs but later rebounds to match the size of controls (Fig. 2B, left).

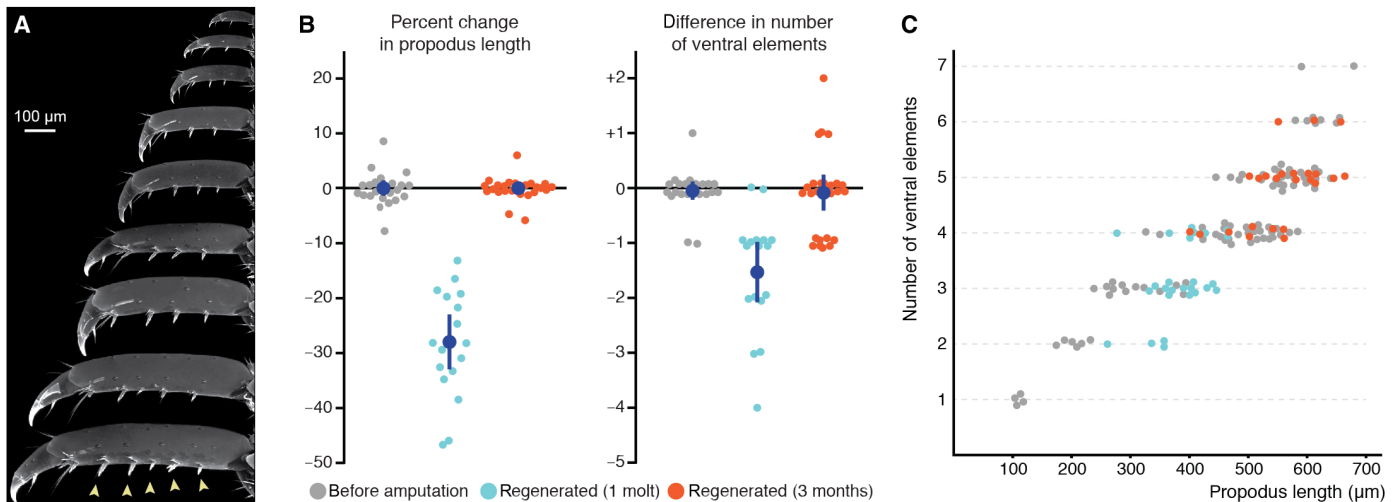
Probing further the relationship between the number of setae and the size of the field in which they develop (see fig. S3C), we found that the number of ventral elements on the propodus increases proportionally with podomere length (Fig. 2, A and C). In the regenerated propodus, the number and the relative spacing of ventral elements matches those of similarly sized uninjured podomeres (Fig. 2C and fig. S3, C and D).

### Sensory organ innervation and function

Immunostainings of acetylated  $\alpha$ -tubulin and a transgenic marker labeling cell outlines revealed that regenerated setae are reinnervated (Fig. 3, A and B, and fig. S2). This is consistent with restoration of their sensory function. To explicitly test this, we assayed

mechanosensory function by touching comb and ventral array setae (Fig. 3A) with a fine probe. This elicited a characteristic escape response (movie S1).

We assessed the sensitivity of *Parhyale* to touch, before and after regeneration, by probing each individual several times and recording the frequency of response (from 0 to 100%) per individual. Tests were conducted separately on three groups of mechanosensory setae that elicit a response rate of 83 to 87% (Fig. 3, A and C, and data file S2). After amputation, two of these groups of setae were eliminated, and responsiveness on the third (the combs of the carpus, which lie just proximally of the amputation plane) dropped to 9 to 18%. After regeneration, response rates on all three groups of setae were indistinguishable from those on control legs of the same individuals (74 to 79%; Fig. 2C). These results show that regenerated legs have the same sensitivity to mechanosensory stimuli as unharmed control legs of the same individuals.



**Fig. 2. Recovery of sensory organ number.** (A) The size of *Parhyale* legs varies with age, shown here for the propodus and dactylus of T4 or T5 legs of increasing age (top to bottom). The ventral arrays are made up of a different number of elements (arrowheads) depending on leg size. (B) Change in podomere size and number of ventral array elements following regeneration. Left: Percent change of propodus length in regenerated versus contralateral uninjured T4 or T5 legs, before amputation (gray), at the first molt after amputation (cyan), and ~3 months after amputation (red). Newly regenerated legs are smaller than uninjured controls, but their size is fully restored 3 months after amputation. Right: Difference in number of ventral array elements in the propodus of the regenerated and contralateral legs. Newly regenerated legs have fewer ventral arrays, but the number is fully restored within 3 months. Mean values and 95% confidence intervals are shown in dark blue. (C) Variation in the number of ventral array elements shown against the length of the propodus in uninjured T4 and T5 legs (gray) and regenerated legs at the first molt after amputation (cyan) and ~3 months after amputation (red). There is a strong linear correlation between podomere size and the number of ventral arrays. At 3 months after amputation, the number of arrays in regenerated legs is the same as in size-matched uninjured legs.

### Diversity and relative abundance of cell types

Considering the faithful recovery of these microanatomical features, we next investigated whether cellular composition and diversity displayed similar high fidelity in regenerated legs. *Parhyale* legs consist of multiple differentiated cell types, including epidermis, muscles, neurons, tendons, glia, and possibly other unknown cells. To characterize this diversity, we established single-nuclei transcriptional profiling of adult *Parhyale* legs [single-nucleus RNA sequencing (snRNAseq)]. This method recovers nuclei from all cell types and overcomes biases in single-cell methods caused by differential dissociation and cell survival, which is especially problematic in animals surrounded by a hard exoskeleton (15–17).

First, we obtained snRNAseq data for 9459 nuclei from T4 and T5 legs of adult *Parhyale* (Fig. 4A). Reads were mapped to the intronic sequences of annotated genes to minimize contamination from ambient cytoplasmic RNA resulting from cell lysis. Dimensionality reduction and clustering analyses separate these nuclei into well-defined clusters, representing cell types or cell states with distinct transcriptional profiles (Fig. 4B, fig. S6, and data file S3). Using known cell-type marker genes, Gene Ontology (GO) analysis, and comparisons with single-cell RNAseq (scRNAseq) profiles obtained in *Drosophila* legs (18), we were able to identify the cell clusters corresponding to epidermis, muscle, neurons, and blood cells (Fig. 4C and figs. S7, A and B, and S8). We also tentatively identified clusters corresponding to glia, trichogen cells, and other support cells that contribute to sensory organs (see Materials and Methods and fig. S7, C and D). The remaining unidentified clusters may correspond to known (e.g., tendons) or previously undescribed cell types in crustacean legs (see data file S3 for cluster-defining transcripts).

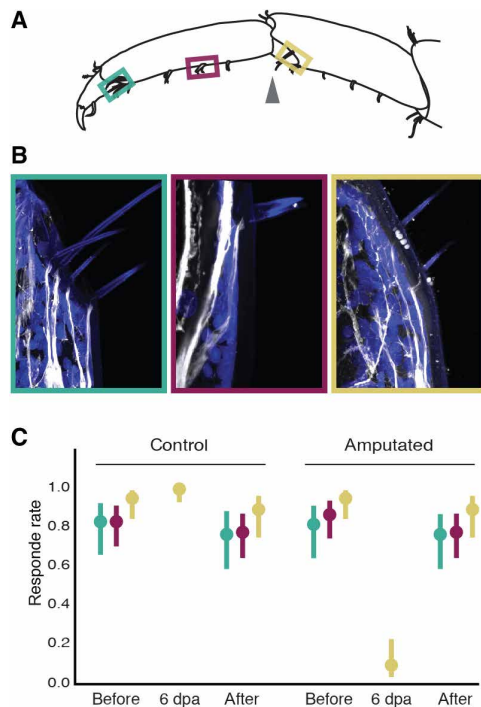
To probe cellular composition after regeneration, we performed snRNAseq on regenerated adult legs approximately 2 and 6 months

following amputation (see fig. S9). To account for genetic, environmental and physiological (e.g., nutritional, circadian, molt-, or age-related) effects on the nuclear transcriptomes, we compared snRNAseq data from regenerated legs (8925 and 6668 nuclei from 2 and 6 months after amputation, respectively) with data from contralateral uninjured legs collected simultaneously from the same individuals (8539 and 7370 nuclei) (Fig. 5A). The analysis was performed de novo on the integrated datasets, projecting the data on a new uniform manifold approximate and projection (UMAP) (Fig. 5, A to C). In both experiments, the regenerated legs have the same cell clusters as their contralateral controls, including all the clusters corresponding to the major identified cell types, the minor clusters, and those corresponding to yet unidentified cell types (Fig. 5, B to D). The top marker genes for each cluster show consistent levels of expression across experiments, in both uninjured and regenerated legs (fig. S10).

We found similar proportions of cells for all clusters in regenerated and contralateral control legs [no significant differences using either a linear model, analysis of variance (ANOVA)  $P$  value of  $>0.08$ , or alternative approaches such as Milo with a false discovery rate (FDR) of  $>10\%$  and scCODA (single-cell compositional data analysis) with an FDR of  $>40\%$  (19, 20); see Materials and Methods]. Power analysis suggests that we would have been able to detect twofold differences in relative abundance for all the major cell clusters (more than half of the clusters, containing ~90% of the cells) and three- to fourfold differences for most of the minor cell clusters (fig. S11).

To probe the recovery of cell types at higher resolution, we investigated whether the scRNAseq data might allow us to distinguish differences in the abundance of specific neuronal cell types (e.g., chemosensory versus mechanosensory neurons). For this, we performed a separate clustering analysis on neuronal cells only. This revealed eight subclusters (fig. S12A) defined by specific expression





**Fig. 3. Recovery of sensory organ innervation and function.** (A) Location of comb and ventral array setae tested. (B) Acetylated  $\alpha$ -tubulin staining of comb and ventral array setae after the first molt after amputation, showing that the setae have been reinnervated. (C) Mechanosensory responses following mechanical stimulation of comb or ventral array setae in control (uninjured) and regenerated legs of the same individuals (assay shown in movie S1). Legs were amputated at the joint of the propodus and the carpus [arrowhead in (A)]. The frequency of response was measured in the same individuals before amputation and after the first molt following regeneration. The comb setae of the propodus (green), the ventral setae on the propodus (magenta), and the comb setae of the carpus (yellow) were tested in separate experiments. The combs of the carpus lie just proximal to the amputation plane and are therefore still present at the distal end of the amputated leg stump, in the region of the blastema. When stimulated 6 days after amputation [6 (days post amputation)], they were found to elicit a very low response rate. Error bars indicate 97.5% confidence intervals.

profiles (fig. S12, A and D, and data file S3), likely representing distinct neuronal cell types. Expression of top markers of *Drosophila* mechanosensory and gustatory neurons does not help to clearly identify the *Parhyale* neural subtypes (fig. S12E). We find that relative abundance of these neuron subtypes is similar in control and regenerated legs (fig. S12, B and C).

To assess the robustness of our analysis, we tested different methods of data integration (21, 22) and different read mapping strategies. In the latter case, we compared results after mapping RNAseq reads either to introns (which should be found in reads confined to the nucleus) or to the entire genome including unannotated regions to account for incomplete gene annotation in the *Parhyale* genome (see Materials and Methods). These analyses gave consistent cell clustering and cell abundance measures in regenerated and control legs (Fig. 5, B to D, and figs. S13 and S14).

Methods for integrating data from separate experiments might fail to distinguish experimental batch effects from real biological variation. To address whether data integration may influence our conclusions, we also analyzed all the datasets by simply aggregating them, without an explicit integration step. Again, we found no

consistent differences in the cell clusters recovered from control and regenerated legs (fig. S15). Within the large epidermal cell cluster, we find that cell abundance among different epidermal subclusters varies by sample rather than by experimental condition (fig. S16). These results show that, even without data integration, regenerated legs are indistinguishable from controls.

In the 6-month experiment, some of our analyses revealed the existence of additional cell clusters in both uninjured and regenerated legs (arrowheads in Fig. 5, A and D, and figs. S13 and S15). These comprise cells expressing molt-associated transcripts ((12) and fig. S17), presumably collected from individuals close to molting. The detection of this molt-associated cell state confirms that our data integration approach is capable of detecting clusters that are present only in some datasets.

### Transcriptional profiles of cell types

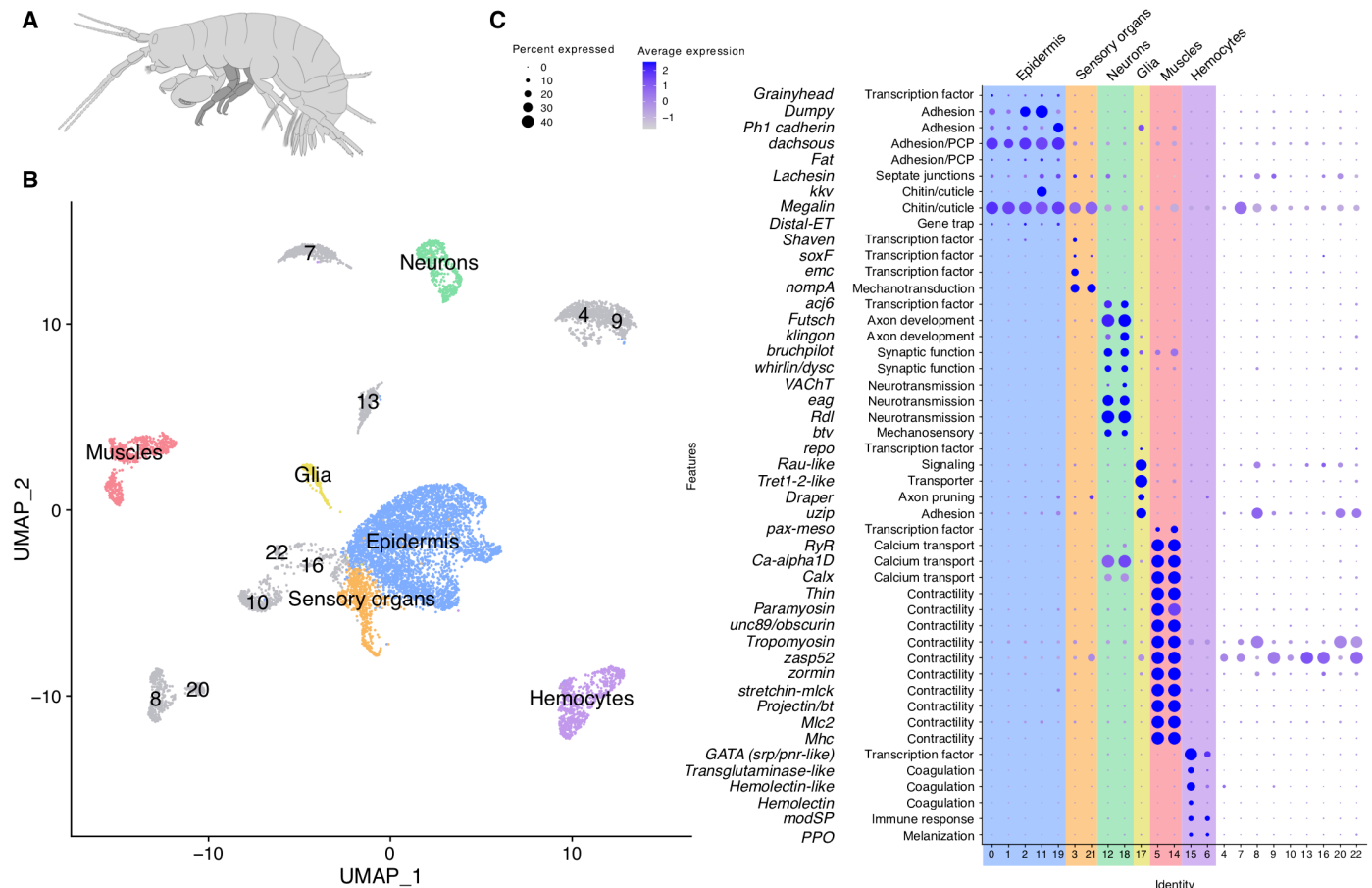
In addition to cell composition, we also compared the transcriptional profile of each cell type between uninjured and regenerated legs by performing differential expression analyses per cell cluster, across all the single-nuclei datasets. We found limited differences, and the great majority represented variation across experiments rather than between regenerated and control legs (fig. S18, B to D). For increased sensitivity, we also performed pseudo-bulk differential expression analysis using DESeq2, with our control and regenerated samples considered as replicates. We found few genes with cluster-specific differential expression in control versus regenerated legs (109 genes with a  $P$  value of  $<0.05$ ), but the great majority of these genes were expressed in a single replicate or in just a handful of cells (fig. S18E).

Last, we used machine learning to test whether differences in these transcriptional profiles would be sufficient to distinguish regenerated from control legs. This approach was able to accurately classify different cell types but could not discriminate between the cells of regenerated and control legs (Fig. 6 and fig. S19).

### DISCUSSION

Overall, our data reveal that the regenerated legs of *Parhyale* are indistinguishable from uninjured legs. This conclusion is based on complementary approaches interrogating the microanatomy, sensory function, cellular composition, and molecular profiles of cells. While we cannot exclude the existence of subtle patterning or functional defects in regenerated *Parhyale* legs, at the current level of morphological and molecular resolution, these legs appear to be perfect replicas of the original legs before amputation.

Our approaches for probing morphology, sensory function, cell diversity, and molecular profiles have some inherent limitations, which are important to acknowledge. On the microanatomical level, although we conclude that leg regeneration in *Parhyale* restores the full pattern and diversity of external sensory organs at the tips of T4 and T5 legs, we have not probed other aspects of morphology and microanatomy, for example, patterns of internal proprioceptive neurons, muscles, tendons, or muscle innervation. Using a behavioral assay, we show that regenerated mechanosensory organs are sensitive to touch, but we are not currently able to probe how sensory signals from these organs become integrated in more complex behaviors. On the cellular level, we are confident that all the identified cell types are recovered during regeneration; if any major cell type was missing [such as the missing bone cells in regenerated



**Fig. 4. Diversity of cell types in *Parhyale* legs revealed by scRNAseq.** (A) Illustration of *P. hawaiiensis* adult, with T4 and T5 legs highlighted in dark gray. (B) UMAP of snRNAseq experiment on uninjured T4 and T5 adult legs depicting 9459 cells clustered based on gene expression. We have identified distinct clusters corresponding to epidermis, muscles, neurons and hemocytes, and putative accessory cells of leg sensory organs and glia (shown in color, see Materials and Methods); the other clusters are labeled by number. (C) Dot plots of selected marker genes for the cell clusters with assigned identities (see table S2). For each marker, we indicate the *Drosophila* or vertebrate ortholog, broad functional category, the average level of expression per cluster (heatmap), and the fraction of cells in each cluster expressing the marker (circle size).

lizard tails; (3, 4)], then our approach would have uncovered that difference. We observe no significant differences in the relative proportions of different cell types between control and regenerated legs, but we acknowledge that with the observed levels of variation between samples, the sensitivity of our approach is limited to detecting >2-fold differences in the abundance of any given cell type. Similarly, although we do not observe significant differences in the molecular profiles of cells in control versus regenerated legs, we recognize that probing differential expression in single-nucleus studies has inherent challenges, for instance, a relatively low sensitivity (23, 24).

The remarkable fidelity that we have observed in *Parhyale* raises interesting questions about the mechanisms underpinning faithful regeneration of complex organs. Previous studies have suggested that the same regulatory networks are deployed to build an organ in embryos and in regenerating adults (25, 26). Recent work in *Parhyale* and in the sea anemone *Nematostella vectensis*, however, shows that development and regeneration have different temporal profiles of gene expression (12, 27). We suggest that development and regeneration could be driven by distinct global regulatory mechanisms yet incorporate some shared genetic modules for patterning and cell differentiation that lead to a conserved outcome.

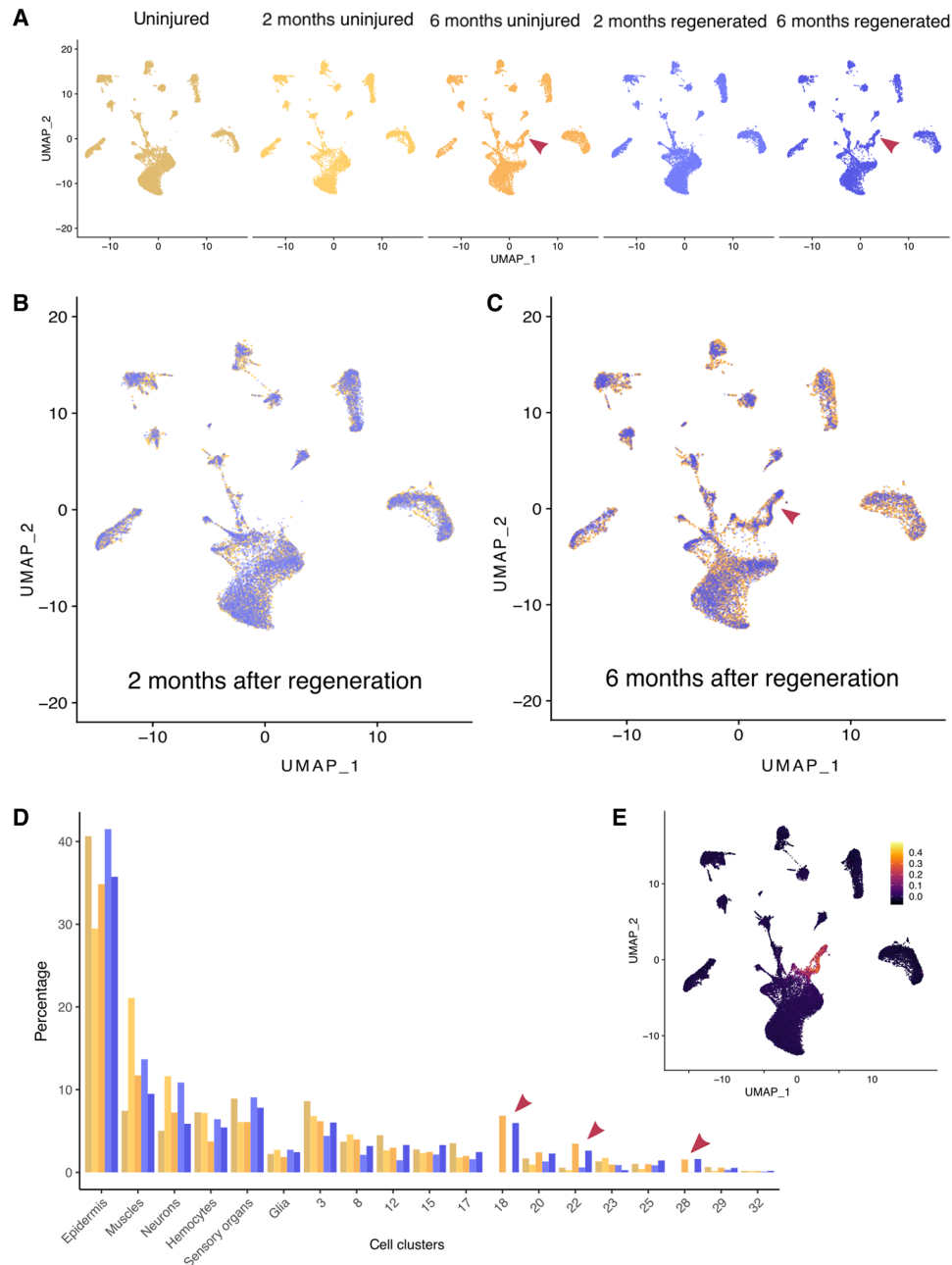
## MATERIALS AND METHODS

### Parhyale culture

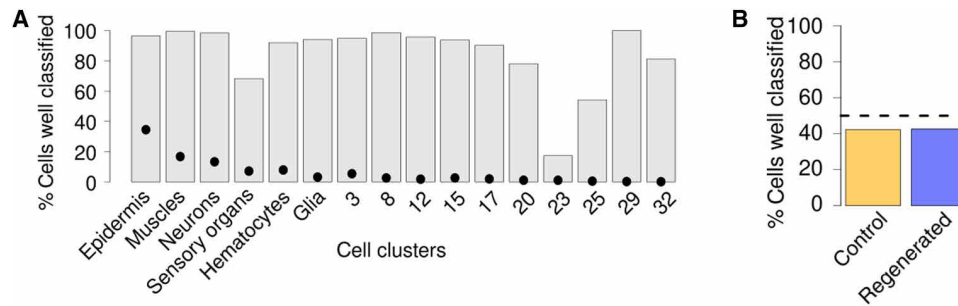
Wild-type *Parhyale* were raised as described previously (28). For all experiments (except fig. S2, see below), we used the wild-type isofemale line Chicago-F (29). Several weeks before snRNAseq experiments, animals were taken from the cultures and kept individually in six-well plates to ensure uniform conditions and to prevent cannibalism. Similarly, before confocal and electron microscopy, animals were kept individually for 3 months in the dark to prevent the buildup of algae. To examine the innervation of setae (fig. S2), we used transgenic animals carrying the *PhHS > lyn-tdTomato-2A-H2B-EGFP* transgene in which lyn-tagged tdTomato (labeling cell outlines) and histone-tagged enhanced green fluorescent protein (EGFP) (labeling nuclei) are expressed under a heat shock promoter (11). Animals were anesthetized using 0.02% clove oil.

### Laser scanning confocal microscopy

For quantification of setae, legs were fixed for 30 min in 3.6% formaldehyde (VWR, #20909.290) in artificial seawater (ASW; specific gravity of 1.02) at room temperature. The legs were then washed for 1 hour in ASW and 30 min in 50% glycerol in phosphate-buffered



**Fig. 5. Recovery of diversity and relative numbers of cell types after regeneration.** (A) Comparison of five snRNAseq datasets (40,961 nuclei in total), including one dataset from uninjured legs (left, same as in Fig. 4B) and datasets from regenerated and control legs sampled ~2 and ~6 months after leg amputation. Samples from uninjured and regenerated legs are highlighted in yellow and blue, respectively. The dimensionality reduction and clustering were performed after integration of the datasets and are independent of those shown in Fig. 4B. (B and C) Overlay of the integrated UMAPs of regenerated and control snRNAseq datasets sampled ~2 months (B) and ~6 months after amputation (C). In (C), the epidermal clusters are affected by the expression of molt-associated transcripts in a subset of cells (arrowhead, see fig. S17). Within each experiment, the same cell clusters are represented in regenerated and control samples. (D) Comparison of the relative number of cells recovered per cluster, across all datasets (expressed as percentage of the total number of cells). Beyond variations between individual experiments, we observe no systematic difference between cell proportions in regenerated versus control legs (depicted in shades of blue and yellow, respectively). Cell clusters 18, 22, and 28 [arrowheads in (A), (C), and (D)] are specifically associated with molting (see main text and fig. S17; see fig. S18A for cluster IDs). (E) Feature plot indicating the average expression of genes found to be associated with molting in *Parhyale* leg transcriptomes (12). These genes are predominantly expressed in the molt-associated clusters [arrowheads in (A), (C), and (D)].



**Fig. 6. Predicting cell type and regeneration status from the transcriptome. (A)** A machine learning approach (see Materials and Methods) can be used to predict each cell's cluster classification reliably for most cell types based on the cell's nuclear transcriptome (frequency of random guess indicated by dots). The lowest prediction accuracy is found for neighboring clusters (for instance, 57% of the cells from cluster 23 are predicted as coming from cluster 8; see figs. S18A and S19C). **(B)** Based on the same machine learning approach, predicting whether a cell derives from a control or a regenerated leg based on its nuclear transcriptome is no better than a random guess (depicted by a dotted line). For further details, see fig. S19.

saline (PBS), transferred to 70% glycerol, and mounted. Cuticle autofluorescence was observed under a Zeiss LSM 800 laser scanning confocal microscope using the 488-nm excitation laser and a 400- to 730-nm detection window. The images shown in Fig. 2A were cropped manually and juxtaposed on a black background.

For the images shown in fig. S2, the T5 leg of a transgenic animal was amputated on the distal part of the carpus. One day after the molt following amputation, the animal was heat-shocked at 37°C for 45 min. One day later, the animal was fixed for 15 min in 3.6% formaldehyde in ASW and washed thoroughly with seawater. The legs were dissected and mounted in VECTASHIELD mounting medium. The native fluorescence of tdTomato and EGFP was imaged on a Zeiss LSM 800 laser scanning confocal microscope.

### Scanning electron microscopy

Adult animals of various sizes were selected, and the T4 and T5 legs on their right side were amputated at the distal part of the carpus (see dashed line, Fig. 1). The uninjured legs on the contralateral side were used as controls. Following the first molt after amputation, the animals were anesthetized, washed in ASW, fixed for 2 hours in 1% glutaraldehyde (Electron Microscopy Sciences, #16300) in ASW, then washed in ASW, and refixed for 2 hours in 1% osmium tetroxide (Electron Microscopy Sciences, #19150) in ASW. Animals were then washed in ASW for 1 hour. All steps were carried out at room temperature.

The regenerated T4 and T5 and their contralateral uninjured legs were carefully removed from the fixed individuals, dehydrated by washing in increasing concentrations of ethanol (in ASW), and stored in 90% ethanol. The samples were handled on the proximal parts of the leg not to damage the distal structures. Subsequently, they were washed three times in absolute ethanol (Merck, #1009832500), subjected to critical point drying in a Leica EM CPD300 critical point dryer, mounted on specimen holders, and coated with gold in a Polaron SC7640 sputter coater. Imaging was performed on a JEOL 6700F scanning electron microscope at the electron microscopy facility of the Stazione Zoologica Anton Dohrn, Naples. The images shown in Fig. 1 (A and C) and fig. S3 were cropped manually and placed on a black background. Color highlights in Fig. 1 and fig. S3 were added as separate layers using the overlay channel option in Adobe Photoshop CS6.

### Description of different types and groups of setae

We used SEM to survey the surface of *Parhyale* legs, focusing on the three most distal podomeres (dactylus, propodus, and carpus) of the T4 and T5 thoracic legs. These two legs show almost identical

patterns of setae. On the leg surface, we observe the following types of setae, which we categorize following nomenclature introduced in previous studies (see also table S1) (14):

1) *Hooked setae* (Fig. 1F and fig. S1B,B'). Simple setae having a long, thin shaft tapering gradually toward the apex. The shaft has a hooked shape, a series of fine nicks before the tapering distal region, and a terminal pore. A single hooked seta is found on the posterior side of the dactylus of T4 and T5 legs.

2) *Curved setae* (Fig. 1E and fig. S1B,B'). Simple setae have long twisting shaft bearing a pore approximately two-thirds along the length of the shaft. A single curved seta is found on the ventral side of the dactylus of T4 and T5 legs.

3) *Plumose setae* (Fig. 1G and fig. S1A,A'). Setae have long shaft bearing two opposed rows of long setules, which give it a feathery appearance. The shaft has no visible pores. A single curved seta is found on the dorsal side of the dactylus of T4 and T5 legs.

4) *Cuspidate setae* (Fig. 1, C and K, and fig. S1C,C'). Simple setae have a relatively short and stout shaft. The shaft bears longitudinal ridges and has no visible pore. A single cuspidate seta is found ventrally on the distal end of the propodus of T4 and T5 legs.

5) *Lamellate type 1 setae* (Fig. 1H and fig. S1E,E'). Lamellate setae consist of a smooth shaft bearing a series of lamellae toward the tip. Lamellate type 1 setae have a wider base and a shorter shaft than lamellate type 2, and they bear a terminal pore. Lamellate type 1 setae are organized in ventral arrays (see below) in the carpus and propodus of T4 and T5 legs.

6) *Lamellate type 2 setae* (Fig. 1I and fig. S1D,D'). Similar to lamellate type 1 setae but tend to have a more slender base, a longer shaft and no pore. Lamellate type 2 setae make up the crown and the anterior and posterior combs (see below) in the carpus and propodus of T4 and T5 legs.

7) *Twin setae* (Fig. 1J and fig. S1E,E'). Twin setae are composites of cuspidate and lamellate setae. The cuspidate-like main shaft bifurcates into a branch that resembles the tip of a typical lamellate seta with a terminal pore. Twin setae usually make up the central seta of each element of the ventral array (see below) in the carpus and propodus of T4 and T5 legs. The relative size of the cuspidate and the lamellate components of twin setae can vary (fig. S4); we suggest that these differences reflect a developmental transition from lamellate type 1 to twin setae.

8) *Microsetae* (Fig. 1, A and D, and fig. S1F,F'). Microsetae are very small setae bearing a terminal pore covered by a hood. One side of the shaft has a lamellated appearance. Microsetae are associated



with characteristic dimples on the cuticular surface (see Fig. 1A). They are dispersed on the surface of the leg, with no stereotypic arrangement.

Lamellate and twin setae are clustered in groups on the propodus and carpus of T4 and T5 legs, which were categorized as follows:

1) Crown group: A row of lamellate type 2 setae located dorsally on the distal end of the propodus and the carpus (marked blue in fig. S3). The crown consists of a variable number of setae (see fig. S5 and data file S1).

2) Comb groups: Two clusters of lamellate type 2 setae located ventrally on the anterior and posterior sides on the distal end of the propodus and the carpus, flanking the first element of the ventral array (marked green in fig. S3). The combs consist of a variable number of setae (see fig. S5 and data file S1); anterior combs are sometimes reduced to a single seta.

3) Ventral array: Groups of regularly spaced lamellate type 1 and twin setae are distributed on the ventral side of the propodus and the carpus (marked yellow in fig. S3). Each element of the array usually consists of a twin seta flanked by lamellate type 1 setae; the element at the distal-most end of the podomere is usually a single lamellate type 1 seta (in juveniles) or a single twin seta (in adults). As suggested earlier, lamellate type 1 setae may mature to become twin setae (fig. S4).

### Testing the spatial distribution of setae

To examine the distribution of ventral array elements in the propodus, we measured the distances between successive elements (as shown in fig. S3C) in control and regenerated legs one molt and 3 months after amputation. The effect of regeneration on this distribution was assessed using a linear mixed-effects model (R function `lme` from the package `nlme` v3.1-157). The individual of origin was tested as a random effect, while the type of leg (T4 or T5), the type of interval measured (proximal, distal, or middle; see fig. S3C), the experimental condition (control, one molt after amputation or 3 months after amputation), or the interaction between the experimental conditions and the type of interval were considered as fixed effects. All factors had an effect. To further identify differences between control and regenerated legs, we used the R function `emmeans` (with parameter “adjust” as “tukey”). We found no significant differences in spacing of setae between control and regenerated legs, except between the most proximal setae and the joint after one molt after amputation ( $P$  value of  $<0.0001$ ; fig. S3D). There was no detectable difference 3 months after amputation.

### Testing mechanosensory function

Large adult males were immobilized on standard 60-mm petri dishes from their body using surgical glue [as described in (11)]. The setae on T4 and T5 legs were stimulated using a minutien pin (Fine Science Tools, #26002-10), and the animals' whole-body response was recorded (movie S1 and data file S2). Two rounds of experiments were performed under dim light; in the first round, the posterior comb setae of the carpus were stimulated twice ( $n = 22$  individuals). After each stimulus, the response was recorded, and the animals were left to relax for 15 min. Following this, either one of the T4 ( $n = 10$ ) or T5 ( $n = 12$ ) legs of each animal was amputated. Six days after amputation, the same setae were stimulated again on both legs, and the animals were left to molt. After molting, they were immobilized again, and the same setae were stimulated on both legs. In the second round, the animals were glued the same way, and the posterior

comb ( $n = 25$ ) and middle ventral array setae ( $n = 28$ ) of the propodus of T4 and T5 legs were stimulated twice with 15-min intervals. After either of the legs was amputated and following the molting, the animals were glued, and the same setae were stimulated.

The effect of amputation and subsequent regeneration on the response to mechanical stimulus was assessed using a generalized linear mixed-effects model (R function `glmer` from the package `lme4` v1.1-27.1) with a binomial distribution of errors. The sample of origin was tested as a random effect, while the leg of origin (T4 or T5), the nature of the leg (amputated or control), and the progress of regeneration (unamputated, 6 days post amputation, regenerated) were considered as fixed effects. The leg (T4/T5) and the sample of origin had no significant effect on response to stimulus. The amputated legs had a significantly lower response rate 6 days after amputation ( $P$  value of  $<10^{-7}$ ), whereas the response in fully regenerated legs was not significantly different from the one in unamputated legs. The error bars in Fig. 3C indicate 97.5% confidence intervals.

### Generation of antibodies and immunostaining

A 746-base pair (bp) fragment from the coding sequence of *Parhyale repo* (gene MSTRG.29918) was amplified by polymerase chain reaction (primers: `aaaatcgataaGATTCAGCACCAGGCA-GAG` and `aaaatcgatCTGCCGAATTCTTCTTGGAC`) and cloned into the *pATH11* plasmid, downstream and in-frame with the *trpE* coding sequence, using the *Cla* I restriction site. The plasmid was transformed to TOP10 *Escherichia coli* competent cells. Transformant colonies were grown in a tryptophan-deficient medium, expression of the TrpE-Repo fusion protein was induced by supplementing the medium with indoleacrylic acid (40  $\mu$ g/ml), and the fusion protein was recovered in inclusion bodies [protocol adapted from (30)]. Mouse immunization and serum collection were performed by Davids Biotechnologie GmbH.

For staining with the Repo antibody (fig. S7, C and D), we collected stage S27 embryos (28) and T4 and T5 regenerating legs 6 days after amputation. For acetylated  $\alpha$ -tubulin staining in Fig. 3B, the legs were collected 1 to 2 days after the first molt following amputation. The animals were anesthetized and fixed at room temperature for 10 min in 3.6% formaldehyde in ASW. The carpus and propodus of regenerated and uninjured control legs were then dissected and cut in half to improve antibody penetration. The cut legs were refixed for 15 min in 3.6% formaldehyde in PBS, washed for 1 hour in PBS with 0.1% Triton X-100 (PTx), and incubated for 1 hour at room temperature in PBS with 0.1% Triton X-100, 0.1% sodium deoxycholate, 0.1% bovine serum albumin (BSA), and 1% normal goat serum (PTxD+). The samples were then incubated for 3 days at 4°C, with mouse monoclonal 6-11B-1 antibody for acetylated  $\alpha$ -tubulin (diluted 1:1000; Sigma-Aldrich, T6793, RRID: AB\_477585) or anti-Repo mouse serum (diluted 1:200, see above) in PTxD+. After washing overnight in PTx at 4°C, the samples were incubated with the secondary antibody (anti-mouse immunoglobulin G Alexa Fluor 488, diluted 1:1000; Life Technologies, A11001, RRID: AB\_2534069) in PTxD+ for 3 days at 4°C. Samples were then washed overnight in PTx at 4°C and mounted in VECTASHIELD mounting medium (Vector Laboratories, H-1000). The samples were imaged on a Zeiss LSM 800 laser scanning confocal microscope.

### Nuclear isolation and sequencing

For each snRNAseq dataset, we collected a pool of 12 to 16 thoracic T4 and T5 legs from large adult males. The smallest specimens had a



body length of ~9 mm (estimated age, >12 months), and the average had a body length of ~11 mm (estimated age, >18 months). The legs were amputated at the proximal end of the basis (excluding the gills and the coxal plates) using a microsurgical knife (Electron Microscopy Sciences, #72047-30) under anesthesia in 0.02% clove oil. To compare regenerated and intact legs, T4 and T5 thoracic legs from one side of the animal were amputated and allowed to regenerate and to grow until ~2 months (8 weeks) after amputation in one pool of individuals or ~6 months (24 weeks) after amputation in a separate pool. The contralateral legs were left intact and used as controls. We then collected the sets of regenerated and control T4 and T5 legs by cutting them off at the basis and dissected them (cut the podomeres lengthwise) in a chilled hypotonic buffer containing 250 mM sucrose, 25 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM tris buffer (pH 8.0), and 1 μM dithiothreitol supplemented with 1× protease inhibitor (Sigma-Aldrich, #11873580001), ribonuclease inhibitor (0.4 U/μl) (Invitrogen, #AM2682), SUPERase-In ribonuclease inhibitor (0.2 U/μl) (Invitrogen, #AM2694), 1 mM spermine, and 0.1 M spermidine [modified from (31)]. The tissue was extruded from the cuticle by gently scraping and peeling using a microsurgical blade and fine forceps (Fine Science Tools, #10316-14 and #11251-20) under a dissecting stereoscope. Triton X-100 was then added to a final concentration of 0.3%, and the tissues were mechanically disrupted by gently pipetting 25 times through a 200-μl filter tip with a cut tip. These steps were carried out in 300 μl of buffer in the lid of a 5-ml DNA LoBind tube (Eppendorf, #0030108302) kept cold on the surface of a frozen metal plate. The disrupted tissue samples were then transferred to a 1.5-ml DNA LoBind Eppendorf tube (Eppendorf, #0030108051), and large pieces of cuticle were let to settle down. The supernatant containing dissociated nuclei was transferred to a new LoBind tube in which 600 μl of fresh hypotonic buffer was added. The nuclei were collected by centrifugation at 500 relative centrifugal force at 4°C for 6 min on a 5424 R Eppendorf microcentrifuge. Centrifugation was kept to a minimum, which resulted in a loose pellet, because longer or faster centrifugations result in lower-quality nuclei. The supernatant was discarded, and the pellet was washed with 600 μl of 1× Dulbecco's phosphate-buffered saline (DPBS) supplemented with 2% BSA. After a second centrifugation, all the supernatant except 25 μl was removed, and an equal volume of 3× DPBS with 2% BSA was added to reach a final concentration of 2× DPBS (we observed that nuclei are better preserved, but they are more difficult to pellet in 2× DPBS). To remove aggregates, the samples were filtered through 10-μm cell strainer (PluriSelect Mini strainer) to generate single-nucleus suspensions. All the steps were performed in the cold, and samples were kept on ice until nucleus capture. The quality of the nuclei was assessed by staining with trypan blue, and nuclei counts were made using a Malassez hemocytometer.

snRNAseq was performed using the 10x Genomics single-cell capturing system. Nuclei suspensions, including 20,000 to 25,000 nuclei, were loaded on the 10x Genomics Chromium Controller following the manufacturer's instructions, within 1 hour of preparation. Single-nucleus complementary DNA (cDNA) libraries were prepared using the Chromium Single Cell 3' Reagent Kit according to the manufacturer's instructions (chemistries v3 and v3.1). Libraries were sequenced on an Illumina NextSeq500 (28 bp for R1 and 122 bp for R2), and 248 to 315 million reads were obtained per library. Reads were mapped to the *P. hawaiiensis* genome version Phaw\_5.0 (see below) using the Cell Ranger Software Suite 6.0.1 [10x Genomics; (32)]. The Cell Ranger output included lists of features, barcodes,

and matrices (corresponding to putative genes, cells, and read counts, respectively). Subsequent analyses were performed using the R package Seurat v4 (21). Quality checks are shown in fig. S20.

### Genome annotation

The snRNAseq reads were mapped on the *P. hawaiiensis* genome assembly Phaw\_5.0 ([https://ncbi.nlm.nih.gov/assembly/GCA\\_001587735.2/](https://ncbi.nlm.nih.gov/assembly/GCA_001587735.2/)) in which 19 poorly assembled mitochondrial scaffolds were removed and replaced by the publicly available complete mitochondrial genome (RefSeq NC\_039402.1). To assign reads to genes, we used the gene annotation described in Sinigaglia *et al.* (12). Many genes in this annotation were of undefined or incorrect strandedness. We defined the correct gene strandedness using information from our snRNAseq data (stranded reads from all five snRNAseq experiments) or by identifying putative coding sequences through sequence similarity. First, we compared the total number of snRNAseq read mapping to opposite strands per gene model. In gene models where at least 80% of reads mapped to one strand (with >20 unambiguously mapped reads in total), we considered that strand as the correct one. In gene models where >20% mapped to opposite strands (or when fewer than 20 reads were recovered), we identified the longest open reading frame on each strand, translated this to a putative protein sequence, and used BLAST (basic local alignment search tool) to search the *Drosophila* and human UniProt databases (UP000000803 and UP000005640, respectively) for similar sequences. Strandedness was then assigned on the basis of the best BLAST score (with a minimum threshold of 50). Using this approach, we were able to assign strandedness to 40,047 of 54,699 gene models (the strandedness of 19,334 genes was modified from the original annotation, 13,990 of which using the snRNAseq data and 5344 of which using coding sequence similarity). For the remaining 14,652 genes, both snRNAseq mapping statistics and BLAST scores were inconclusive, and therefore, both strands were included in the gene models (none of those genes had introns). This analysis gave a final list of 69,351 gene models.

As a significant fraction of snRNAseq reads do not map to the exons and introns of the annotated gene models (~40% of reads), we hypothesized that additional genes could be present in the unannotated part of the genome. To maximize read mapping and to examine whether nonannotated genes contribute to the snRNAseq signal, we built a secondary annotation that covers the unannotated part of the genome. For this, we split the intergenic regions of the entire genome into nonoverlapping 5-kb fragments (removing fragments below 1 kb, located near genes) and kept the fragments in which we could detect at least 20 unambiguously mapped snRNAseq reads. Strandedness was assigned to each fragment based on the snRNAseq reads mapping to each fragment, as described above (strand assigned if at least 80% of the reads mapped to one strand, and both strands included if >20% of the reads mapped to opposite strands). The resulting set included 452,770 annotated intergenic features.

### Analysis of snRNAseq data

The analysis was performed using the Seurat v4.0.3 scRNA-seq R package, following the associated documentation (21). Genes transcribed in fewer than five cells were removed from the analysis. Cells with fewer than 100 transcribed genes or more than 10% of the genes mapped to the mitochondrial genes were also excluded from the analysis. Datasets were normalized (Seurat function LogNormalize), and variable genes were found using the variance-stabilizing transformation

(vst) method with a maximum of 2000 variable features. Next, the datasets were scaled, and principal components analysis (PCA) was performed. A shared nearest neighbor (SNN) graph was computed with 20 dimensions (resolution 1) to identify the clusters. A UMAP was used to perform clustering and dimensionality reduction. Known markers for epidermis, neurons, sensory organs, glia, muscles, and hemocytes were plotted using the Seurat function DotPlot (Fig. 4C). We identified differentially expressed (DE) genes in each cluster by comparing the expression profiles of genes between the cells or a cluster and all the other cells using Wilcoxon rank sum–based approach, with a log fold change of  $>0.25$  and a minimum cell percentage of 0.1. The resulting list of DE genes for each cell type was further annotated by identifying putative gene orthologs from *Drosophila melanogaster* and *Homo sapiens* based on reciprocal best BLASTP (protein BLAST) hits. Selected markers among DE genes were displayed individually using the Seurat function FeaturePlot (fig. S6). The top five DE markers per cluster were selected to generate a dot plot color coded by dataset using the package ggplot2 version 3.3 (fig. S10).

We functionally annotated our cell clusters by GO term enrichment analysis using the topGO v2.40.0 package (A. Alexa and J. Rahnenfuhrer, 2021, topGO: Enrichment Analysis for Gene Ontology. R package version 2.48.0. DOI: 10.18129/B9.bioc.topGO) and the database org.Dm.eg.db v3.11.4 (M. Carlson, 2019. org.Mm.eg.db: Genome wide annotation for mouse. R package version 3.8.2. DOI: 10.18129/B9.bioc.org.Mm.eg.db) on the list of DE genes.

To compare the *Parhyale* cell clusters with cell clusters identified in *Drosophila* legs, we first identified the sets of genes that are differentially in each *Drosophila* leg cluster (18) using default parameters in Seurat. Then, we used the putative orthologs of those genes in *Parhyale* (as described above) to score our nuclei using the Seurat function AddModuleScore. The results were displayed using the Seurat function DotPlot (fig. S8).

To integrate and compare the five snRNAseq datasets, we used the anchoring approach provided by Seurat. First, we extracted the raw counts from each Seurat object, performed data normalization (function LogNormalize), and found variable features independently for each dataset. Then, we performed the identification and integration of the anchors with 20 dimensions, followed by scaling, PCA, and the construction of an SNN graph with 20 dimensions (resolution 1). A UMAP was used to display clustering and dimensionality reduction. Closely related cell clusters were merged and labeled on the basis of the cell-type markers shown in Fig. 4C. The relative numbers of cells assigned to each cluster were extracted, calculated from the Seurat object, and visualized using the function barplot of the ggplot2 v3.3.6 R package.

We also repeated this analysis on Seurat using nonintegrated data to test the effect of the anchoring method on the clustering and the recovery of the putative cell types in each dataset (we obtained similar results, shown in fig. S15). To explore potential differences in cell distribution within a complex cluster, we examined more specifically cell abundance and differential expression of top markers across the subclusters that comprise the epidermal cluster (fig. S16).

To examine the impact of the gene annotation in the analysis, we also performed the same analysis as described above using the entire genome, including parts of the genome that lack gene annotations (see the “Genome annotation” section), for read mapping. The two integrated datasets were compared using Clustree v0.4.3 (33) to visualize the overlap of clusters. The datasets were also integrated

using Liger v1.0.0 package (22) and SeuratWrappers v0.3.0 to explore the effect of the integration method. The datasets were first pre-processed separately (normalization, finding most variable features, and scaling the data without centering) according to the Liger tutorial. We then calculated the integrative nonnegative matrix factorization with 20 matrix factors to identify the joint clusters and perform quantile normalization by dataset, factor, and cluster to integrate the datasets. In addition, we ran Louvain community detection (resolution 1) and UMAP on Liger to visualize the clustering of cells graphically. The approach described above was used to merge related clusters.

To probe the cell clusters at higher resolution, we investigated whether the scRNAseq data might allow us to distinguish differences in the abundance of specific neuronal subtypes. Subclustering analysis was performed by selecting the neural cells (clusters 12 and 18 in Fig. 4C) and performing a similar analysis as described above. We extracted the raw counts from the Seurat integrated object, performed data normalization (function LogNormalize), and found variable features independently for each snRNAseq dataset. Then, we performed the identification and integration of the anchors with 30 dimensions, followed by scaling, PCA, and the construction of an SNN graph with 30 dimensions (resolution 1). To counteract the oversplitting of clusters, we applied an approach to merge cluster pairs that differed by less than 20 DE genes with twofold change (34). This process was performed iteratively until all clusters were sufficiently distinct. We thus identified eight distinct neural subclusters (fig. S12). The DE genes for each cluster were identified, and putative *Drosophila* orthologs were annotated as described above. We probed the expression of DE markers in mechanosensory and gustatory receptor neurons of *Drosophila* in these subclusters, as described above (fig. S12E).

### Identification of cell types in the snRNAseq data

Cluster analysis on our snRNAseq data identified approximately 23 distinct cell clusters. To identify the cell types represented in these clusters, we took three complementary approaches. First, we identified transcripts that are enriched within each cluster and performed a GO analysis (fig. S7, A and B). Second, we examined the expression of selected marker genes, whose expression and function are associated with specific cell types throughout the metazoa, or in *Drosophila*, which is the closest extensively studied model organism to *Parhyale* (table S2). Third, in one case, it was possible to use antibodies to observe the localization of a cluster-specific marker in the legs of *Parhyale* (fig. S7, C and D).

Because of intrinsic limitations, we expected that droplet-based snRNAseq would detect only a small subset of all the mRNAs present in a tissue and to capture different mRNAs with different efficiency. This is, in part, because nuclear preparations only capture mRNAs in the transitory phase before their export into the cytoplasm. The residence time of each mRNA in the nucleus varies, e.g., depending on the size of the transcription unit (time to transcribe). A second reason for this differential capture is that standard droplet-based approaches rely on the use of oligo(dT) primers to capture the mRNAs and to initiate cDNA synthesis. Nascent transcripts present in the nucleus have not yet acquired polyadenylated tails, so the oligo(dT) primers usually work by binding spurious stretches of adenosines along the length of the primary transcript (sometimes within long introns). Thus, the efficiency of detecting a given mRNA depends on its residence time in the nucleus (related to the length

of the transcription unit) and the presence of these fortuitous priming sites. Despite these limitations, a given transcript should be detected with similar efficiency across all cell types. Therefore, droplet-based snRNAseq allows us to identify cell types/states based on the differential expression of the subset of mRNAs that are efficiently captured.

We focus here on marker transcripts that could be found in our datasets. Orthology assignments are based on reciprocal best BLAST hits with *Drosophila* [for gene names, see FlyBase (<http://flybase.org/>)]. The numbering of cell clusters below refers to the analysis (UMAP and dot plots) shown in Fig. 4.

#### Epidermal cells

Cell clusters 0, 1, 2, 11, and 19 form a large supercluster of cells (almost 40% of captured cells) expressing overlapping sets of markers. Genes whose functions are associated with the arthropod exoskeleton [chitin synthase *kkv*; (35)], epithelial adhesion to the exoskeleton [*dumpy*; (36)], adherens junctions [Ph1 cadherin; (37, 38)], planar cell polarity [*fat* and *dachsous*; (39)], septate junctions [*Lachesin*; (40)], and a transcription factor associated with epidermal repair [*grainyhead*; (41)] are specifically expressed in this supercluster (see Fig. 4C). We also find that a gene corresponding to an exon trap expressed in the epidermis of *Parhyale* limbs [*Distal-ET*; (42)] is expressed in this supercluster. Individual clusters may represent different subtypes of epidermal cells (e.g., associated with different parts of the leg) or cell states.

#### Muscles

GO analysis provides strong evidence for the identification of clusters 5 and 14 as muscle; GO terms for muscle development and differentiation, mesoderm development, calcium transport and sarcomere assembly are strongly enriched (fig. S7A). Genes with conserved muscle-specific functions—such as genes associated with muscle development [*pox meso*; (43)], contractility [e.g., *Mhc*, *projectin/bt*, and *tropomyosin*; (44)], and calcium ion transport [*Calx*, *Ca-alpha1D*, and *RyR*; (45)]—are most strongly expressed in these clusters (see Fig. 4C).

#### Neurons

GO analysis provides strong evidence for the identification of clusters 12 and 18 as neurons; GO terms for synaptic transmission and membrane potential are strongly enriched (fig. S7B). Genes with neuron-specific functions—including genes associated with neuronal cell type specification [*acj6*; (46)], the formation of axons and dendrites [*futsch*; (47)], synaptic function [*bruchpilot* and *whirlin/dysc*; (48, 49)], mechanosensory function [*btrv*; (50)], and neurotransmission [*VACHT*, *eag*, and *Rdl*; (51–53)]—are most strongly expressed in these clusters (see Fig. 4C). In *Drosophila*, these genes are associated with mechanosensory and/or chemosensory neurons, suggesting that this set is likely to include different types of sensory neurons.

#### Hemocytes

Hemocytes (blood cells) enter *Parhyale* legs through the blood circulation. Cells clusters 6 and 15 express genes that are typically associated with arthropod hemocytes, encoding proteins involved in blood coagulation [related to hemolectin and transglutaminase; (54, 55)], melanization [phenoloxidase (PPO); (56)], and immune responses [modSP; (57)] (see Fig. 4C). We find that these cells also express specifically a GATA transcription factor and a *mys/itgbn*-like  $\beta$ -integrin, whose orthologies could not be precisely determined; GATA factors are typically associated with the specification of endodermal tissues and blood (58), and the  $\beta$ -integrins *mys* and

*itgbn* play essential roles in encapsulation, hemocyte migration, and phagocytosis in *Drosophila* (59–61).

#### Putative glia

Cell cluster 17 is associated with the expression of the putative ortholog of the insect glial cell marker *repo* (62, 63). Using an antibody that we raised against this transcription factor, we find that these cells are closely associated with axons (fig. S7, C and D), as would be expected of glia. The cluster is also enriched in the expression of genes associated with signaling, adhesion, sugar transport, and axon pruning in the context of neuron-glia interactions in *Drosophila* [*rau*, *uzip*, *Tret1-2*, and *draper*; (64–68)].

#### Putative sensory organ accessory cells

Cell clusters 3 and 21 are associated with the large epidermal supercluster described above. These clusters are enriched for expression of a few genes, which are associated with three types of sensory organ accessory cells in *Drosophila* (see Fig. 4C): *soxF/sox15* and *shaven/Pax2*, encoding transcription factors that play essential roles in the differentiation of sensory organ socket cells (69) and shaft cells (70), respectively; and *nompA*, which plays an essential role in the sheath cells of mechanosensory organs (71).

#### Putative sensory organ trichogen (shaft) cells

Cell cluster 7 is associated with expression of the orthologs of spineless, encoding a transcription factor associated with sensory organs in *Drosophila* (72), and hook, encoding a protein of the endocytic pathway that participates in the formation of sensory bristles by the trichogen cells (73).

#### Genes and cell clusters associated with molting

Using an existing dataset (12), we identified 1294 genes that are strongly up-regulated in *Parhyale* legs 3 to 4 days before molting [DESeq2;  $P_{\text{adj}} < 0.001$ ; (74)]. We used the Seurat function `AddModuleScore` to measure the average expression of these genes in our nuclei, and the results were displayed using the Seurat function `FeaturePlot` (Fig. 5E). We used the Seurat function `Dotplot` to compare the expression of top 57 up-regulated genes between cell clusters (fig. S17).

#### Comparing the distribution of cells across cell clusters

We tested whether the proportion of cells recovered in each cell cluster is similar in controls and regenerated legs. We used ANOVA to test the effect of each of three factors on cell count: which condition each cell is associated with (control/regenerated), which cell cluster it belongs to, and which experiment it was sampled from (0/2/6 months; all experiments shown in Fig. 5A). For the regression, we used a negative binomial distribution because it captured best the dispersion of our cell count dataset. We used the `Anova` function from the R package `car` with an F-test and deviance as error estimate.

We found that there was no significant effect of the control/regeneration condition (either as a simple explanatory variable or in interaction with other variables). However, the experiment and the cell type of origin had a strong impact on cell count distribution ( $P$  values of  $<0.001$  and  $<2.2 \times 10^{-16}$ , respectively).

To further identify differences between experiments and cell clusters, we used the R function `emmeans`. In the 6-month dataset, there were significantly more cells in one of the cell clusters associated with molting (cluster 18 in Fig. 5D and cluster 12 in fig. S13B). In addition, there were significantly more epidermal cells in the 0-month dataset (compared to the 6-month dataset) and significantly more cells in the muscle and neuronal clusters in the 2-month



dataset compared to 0- or 6-month dataset. However, no statistically significant difference was observed between cells from regenerated legs and controls in any cluster. When mapping reads to the entire genome, epidermal cells are more abundant after regeneration ( $P$  value of  $\sim 0.01$ ), but this difference disappears if we consider the cluster associated with molting (cluster 12) as part of the epidermal cluster.

We also tested two additional methods: Milo (19) and scCODA (20). With the miloR v1.2.0 package, we followed the instructions given in the vignette with default parameters (an additional set of parameters used in the vignette was also tested with same results: buildGraph function was run with parameters  $k = 30$ ,  $d = 30$ , and reduced.dim = "pca.corrected" and makeNhoods function was run with parameters prop = 0.1,  $k = 30$ ,  $d = 30$ , and refined = TRUE). No significant differences in cell abundance were detected between control and regenerated conditions (FDR = 10%), whereas the batch effect (set of animals sampled) could be detected. With the Python program scCODA v0.1.7, we followed the instructions of the authors. No difference in cell abundance was detected between control and regenerated conditions (FDR = 40%).

### Power of detection in comparisons of cell abundance

To quantify our power to detect differences in cell abundance between control and regenerated samples, we simulated datasets with the same structure as the original dataset, except that we lowered the number of cells of a given cluster in the regenerated samples by 1.2- to 10-fold compared to the corresponding control sample. We then used an ANOVA test (same as described above) to determine whether we would detect the change as significant. We reported the minimum fold change for which a difference in cell abundance was significantly detected ( $P$  value of  $< 0.05$ ; fig. S11). This analysis was repeated for each cluster.

### Differential expression analysis

Differential expression analysis of uninjured versus regenerated legs was performed separately on nonnormalized and nonintegrated counts of the 2- and 6-month datasets using the R package Seurat v4.0.3 function FindMarkers and the bimod method. We chose this method because it gave the most exhaustive list of DE genes compared to other methods (Wilcoxon test,  $t$  test, logistic regression test, and the MAST package), which gave very similar but smaller lists of DE genes. Visualization of gene abundance for the DE genes (in fig. S18D) was done using the Seurat function DotPlot.

For differential expression on pseudo-bulk data, counts were aggregated for each cell type using the Seurat function AggregateExpression, and further analyses were performed using DESeq2 v1.36.0. Genes with fewer than five counts in total were filtered out. The three control and two regenerated samples were considered as two sets of replicates. Genes that were DE in a cluster-specific manner between regenerated and control samples were identified by grouping the factors "cond" (control/regenerated) and "cluster" and taking batch effect (set of animals sampled) into account (formula: "~ batch + group"). One hundred nine DE genes were identified ( $P$  value of  $< 0.05$ ). These genes were detected in very few cells and/or DE in one replicate only (see fig. S18E), with the exception of three genes: MSTRG.32192 [orthologous to *Drosophila* collagen alpha-1(IV) chain], MSTRG.36270 (orthologous to RNA binding protein Rbfox1, also identified as DE in the analysis at the single-cell level), and MSTRG.35733 (orthologous to RluA-1). These genes are mildly enriched in the glia cluster of regenerated samples.

### Prediction of control versus regenerated condition based on transcriptome

To test the presence of a transcriptional signature of regeneration, we tried to predict the regeneration state (regenerated versus control) for each cell based on our snRNAseq data using the R package scPred v1.9.2 (75). scPred was designed to predict cluster of origin, but we reasoned that it could be used to predict any given status of a cell. As a control, we also used the same procedure to predict the cluster of origin of each cell (epidermis, muscle, etc.).

The 2- and 6-month datasets were subsampled, such that the number of uninjured and regenerated cells was identical within each cluster for each experiment: We sampled 7291 cells for each of the uninjured and regenerated 2-month experiments (a total of 14,582 cells) and 5694 cells for each of the uninjured and regenerated 6-month experiments (a total of 11,388 cells).

Each dataset was normalized and scaled, and a PCA was computed using the Seurat functions NormalizeData, ScaleData, and RunPCA.

Training of the machine learning model was done either on the 6-month or the 2-month datasets using the scPred functions getFeatureSpace and trainModel with the model svmRadial (Fig. 6 and fig. S19, D and E, respectively). We tested other models (e.g., mda, knn, and random forest), but they gave either similar or worse results on the training set.

Testing was done on the dataset that had not been used for training, measuring the percentage of cells for which a correct prediction was made by the trained model. Baseline values corresponding to the percentage of cells that would be correctly assigned just by chance were calculated as the proportions of cells present in the dataset for each category (for instance, 50 and 36% of cells in the 2-month dataset classified as "regenerated" or "epidermal," respectively).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abn9823>

### REFERENCES AND NOTES

1. T. H. Morgan, *Regeneration* (The Macmillan Company, 1901).
2. C. W. Beck, J. C. Izpisua Belmonte, B. Christen, Beyond early development: *Xenopus* as an emerging model for the study of regenerative mechanisms. *Dev. Dyn.* **238**, 1226–1248 (2009).
3. R. J. Goss, *Principles of Regeneration* (Academic Press, 1969).
4. K. E. McLean, M. K. Vickaryous, A novel amniote model of epimorphic regeneration: The leopard gecko, *Eublepharis macularius*. *BMC Dev. Biol.* **11**, 50 (2011).
5. C. Kaars, S. Greenblatt, C. R. Fournier, Patterned regeneration of internal femoral structures in the cockroach, *Periplaneta americana* L. *J. Exp. Zool.* **230**, 141–144 (1984).
6. G. Lin, Y. Chen, J. M. W. Slack, Regeneration of neural crest derivatives in the *Xenopus* tadpole tail. *BMC Dev. Biol.* **7**, 56 (2007).
7. R. Diogo, E. Nacu, E. M. Tanaka, Is salamander limb regeneration really perfect? Anatomical and morphogenetic analysis of forelimb muscle regeneration in GFP-transgenic axolotls as a basis for regenerative, developmental, and evolutionary studies. *Anat. Rec.* **297**, 1076–1089 (2014).
8. A. S. Azevedo, S. Sousa, A. Jacinto, L. Saúde, An amputation resets positional information to a proximal identity in the regenerating zebrafish caudal fin. *BMC Dev. Biol.* **12**, 24 (2012).
9. A. Bertozzi, C.-C. Wu, P. D. Nguyen, M. D. Vasudevarao, M. A. Mulaw, C. D. Koopman, T. P. de Boer, J. Bakkers, G. Weidinger, Is zebrafish heart regeneration "complete"? Lineage-restricted cardiomyocytes proliferate to pre-injury numbers but some fail to differentiate in fibrotic hearts. *Dev. Biol.* **471**, 106–118 (2020).
10. N. Konstantinides, M. Averof, A common cellular basis for muscle regeneration in arthropods and vertebrates. *Science* **343**, 788–791 (2014).
11. F. Alwes, C. Enjolras, M. Averof, Live imaging reveals the progenitors and cell dynamics of limb regeneration. *eLife* **5**, e19766 (2016).
12. C. Sinigaglia, A. Almazán, M. Lebel, M. Sémon, B. Gillet, S. Hughes, E. Edsinger, M. Averof, M. Paris, Distinct gene expression dynamics in developing and regenerating limbs. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2119297119 (2022).



13. V. Hartenstein, Development of insect sensilla, in *Comprehensive Molecular Insect Science*, L. I. Gilbert, K. Iatrou, S. S. Gill, Eds. (Elsevier, 2005), pp. 379–419.
14. A. Garm, L. Watling, in *Functional Morphology and Diversity* (Oxford, 2013), pp. 167–198.
15. R. V. Grindberg, J. L. Yee-Greenbaum, M. J. McConnell, M. Novotny, A. L. O'Shaughnessy, G. M. Lambert, M. J. Araúzo-Bravo, J. Lee, M. Fishman, G. E. Robbins, X. Lin, P. Venepally, J. H. Badger, D. W. Galbraith, F. H. Gage, R. S. Lasken, RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19802–19807 (2013).
16. B. B. Lake, S. Codeluppi, Y. C. Yung, D. Gao, J. Chun, P. V. Kharchenko, S. Linnarsson, K. Zhang, A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
17. H. Wu, Y. Kiritani, E. L. Donnelly, B. D. Humphreys, Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: Rare cell types and novel cell states revealed in fibrosis. *J. Am. Soc. Nephrol.* **30**, 23–32 (2019).
18. H. Li, J. Janssens, M. De Waegeneer, S. S. Kolluru, K. Davie, V. Gardeux, W. Saelens, F. P. A. David, M. Brbić, K. Spanier, J. Leskovec, N. P. McLaughlin, Q. Xie, R. C. Jones, K. Brueckner, J. Shim, S. G. Tattikota, F. Schnorrer, K. Rust, T. G. Nystul, Z. Carvalho-Santos, C. Ribeiro, S. Pal, S. Mahadevaraju, T. M. Przytycka, A. M. Allen, S. F. Goodwin, C. W. Berry, M. T. Fuller, H. White-Cooper, E. L. Matunis, S. D. Nardo, A. Galenza, L. E. O'Brien, J. A. T. Dow; FCA Consortium, H. Jasper, B. Oliver, N. Perrimon, B. Deplanck, S. R. Quake, L. Luo, S. Aerts, D. Agarwal, Y. Ahmed-Braimah, M. Arbeitman, M. M. Ariss, J. Augsburg, K. Ayush, C. C. Baker, T. Banisch, K. Birker, R. Bodmer, B. Bolival, S. E. Brantley, J. A. Brill, N. C. Brown, N. A. Buehner, X. T. Cai, R. Cardoso-Figueiredo, F. Casares, A. Chang, T. R. Clandinin, S. Crasta, C. Despland, A. M. Detweiler, D. B. Dhakan, E. Donà, S. Engert, S. Floc'hlay, N. George, A. J. González-Segarra, A. K. Groves, S. Gumbin, Y. Guo, D. E. Harris, Y. Heifetz, S. L. Holtz, F. Horns, B. Hudry, R.-J. Hung, Y. N. Jan, J. S. Jaszczak, G. S. X. E. Jefferis, J. Karkanas, T. L. Karr, N. S. Katheder, J. Kezos, A. A. Kim, S. K. Kim, L. Kockel, N. Konstantinides, T. B. Kornberg, H. M. Krause, A. T. Labott, M. Laturney, R. Lehmann, S. Leinwand, J. Li, J. S. S. Li, K. Li, K. Li, L. Li, T. Li, M. Litovchenko, H.-H. Liu, Y. Liu, T.-C. Lu, J. Manning, A. Mase, M. Matera-Vatnick, N. R. Matias, C. E. McDonough-Goldstein, A. M. Geever, A. D. McLachlan, P. Moreno-Roman, N. Neff, M. Neville, S. Ngo, T. Nielsen, C. E. O'Brien, D. Osumi-Sutherland, M. N. Özel, I. Papatheodorou, M. Petkovic, C. Pilgrim, A. O. Pisco, C. Reisenman, E. N. Sanders, G. D. Santos, K. Scott, A. Sherlekar, P. Shiu, D. Sims, R. V. Sit, M. Slaidina, H. E. Smith, G. Sterne, Y.-H. Su, D. Sutton, M. Tamayo, M. Tan, I. Tastekin, C. Treiber, D. Vacek, G. Vogler, S. Waddell, W. Wang, R. I. Wilson, M. F. Wolfner, Y.-C. E. Wong, A. Xie, J. Xu, S. Yamamoto, J. Yan, Z. Yao, K. Yoda, R. Zhu, R. P. Zinzen, Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
19. E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, J. C. Marioni, Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
20. M. Büttner, J. Ostner, C. L. Müller, F. J. Theis, B. Schubert, scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
21. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, C. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
22. J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, E. Z. Macosko, Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
23. E. Papalexi, R. Satija, Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).
24. J. W. Squair, M. Gautier, C. Kathe, M. A. Anderson, N. D. James, T. H. Hutson, R. Hudelle, T. Qaiser, K. J. E. Matson, Q. Barraud, A. J. Levine, G. L. Manno, M. A. Skinnider, G. Courtine, Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
25. K. Muneoka, S. V. Bryant, Evidence that patterning mechanisms in developing and regenerating limbs are the same. *Nature* **298**, 369–371 (1982).
26. E. Nacu, E. M. Tanaka, Limb regeneration: A new development? *Annu. Rev. Cell Dev. Biol.* **27**, 409–440 (2011).
27. H. Johnston, J. F. Warner, A. R. Amiel, K. Nedoncelle, J. E. Carvalho, E. Röttinger, Whole body regeneration deploys a rewired embryonic gene regulatory network logic. bioRxiv 658930 (2021), <https://doi.org/10.1101/658930>.
28. W. E. Browne, A. L. Price, M. Gerberding, N. H. Patel, Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. *Genesis* **42**, 124–149 (2005).
29. D. Kao, A. G. Lai, E. Stamatakis, S. Rosic, N. Konstantinides, E. Jarvis, A. di Donfrancesco, N. Pouchkina-Stancheva, M. Sémon, M. Grillo, H. Bruce, S. Kumar, I. Siwanowicz, A. Ie, A. Lemire, M. B. Eisen, C. Extavou, W. E. Browne, C. Wolff, M. Averof, N. H. Patel, P. Sarkies, A. Pavlopoulos, A. Aboobaker, The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **5**, e20062 (2016).
30. T. Hoey, Expression and purification of lacZ and trpE fusion proteins. *Curr. Protoc. Mol. Biol.* **Chapter 16**, Unit16.5 (2001).
31. S. R. Krishnaswami, R. V. Grindberg, M. Novotny, P. Venepally, B. Lacar, K. Bhutani, S. B. Linker, S. Pham, J. A. Erwin, J. A. Miller, R. Hodge, J. K. McCarthy, M. Kelder, J. McCorrison, B. D. Aeversmann, F. D. Fuentes, R. H. Scheuermann, J. Lee, E. S. Lein, N. Schork, M. J. McConnell, F. H. Gage, R. S. Lasken, Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).
32. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. M. Farland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. H. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
33. L. Zappia, A. Oshlack, Clustering trees: A visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, gij083 (2018).
34. J. M. Musser, K. J. Schippers, M. Nickel, G. Mizzon, A. B. Kohn, C. Pape, P. Ronchi, N. Papadopoulos, A. J. Tarashansky, J. U. Hammel, F. Wolf, C. Liang, A. Hernández-Plaza, C. P. Cantalapiedra, K. Achim, N. L. Schieber, L. Pan, F. Ruperti, W. R. Francis, S. Vargas, S. Kling, M. Renkert, M. Polikarpov, G. Bourenkov, R. Feuda, I. Gaspar, P. Burkhardt, B. Wang, P. Bork, M. Beck, T. R. Schneider, A. Kreshuk, G. Wörheide, J. Huerta-Cepas, Y. Schwab, L. L. Moroz, D. Arendt, Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science* **374**, 717–723 (2021).
35. S. Ostrowski, H. A. Dierick, A. Bejsovec, Genetic control of cuticle formation during embryonic development of *Drosophila melanogaster*. *Genetics* **161**, 171–182 (2002).
36. M. B. Wilkin, M. N. Becker, D. Mulvey, I. Phan, A. Chao, K. Cooper, H. J. Chung, I. D. Campbell, M. Baron, R. MacIntyre, *Drosophila dumpy* is a gigantic extracellular protein required to maintain tension at epidermal-cuticle attachment sites. *Curr. Biol.* **10**, 559–567 (2000).
37. H. Oda, K. Tagawa, Y. Akiyama-Oda, Diversification of epithelial adherens junctions with independent reductive changes in cadherin form: Identification of potential molecular synapomorphies among bilaterians. *Evol. Dev.* **7**, 376–389 (2005).
38. M. Sasaki, Y. Akiyama-Oda, H. Oda, Evolutionary origin of type IV classical cadherins in arthropods. *BMC Evol. Biol.* **17**, 142 (2017).
39. H. Strutt, D. Strutt, How do the Fat-Dachsous and core planar polarity pathways act together and independently to coordinate polarized cell behaviours? *Open Biol.* **11**, 200356 (2021).
40. M. Llimargas, M. Strigini, M. Katidou, D. Karagozeos, J. Casanova, Lachesin is a component of a septate junction-based mechanism that controls tube size and epithelial integrity in the *Drosophila* tracheal system. *Development* **131**, 181–190 (2004).
41. K. A. Mace, J. C. Pearson, W. McGinnis, An epidermal barrier wound repair pathway in *Drosophila* is mediated by grainy head. *Science* **308**, 381–385 (2005).
42. Z. Kontarakis, A. Pavlopoulos, A. Kiupakis, N. Konstantinides, V. Douris, M. Averof, A versatile strategy for gene trapping and trap conversion in emerging model organisms. *Development* **138**, 2625–2630 (2011).
43. H. Duan, C. Zhang, J. Chen, H. Sink, E. Frei, M. Noll, A key role of Pox meso in somatic myogenesis of *Drosophila*. *Development* **134**, 3985–3997 (2007).
44. F. Schnorre, C. Schönbauer, C. C. H. Langer, G. Dietzl, M. Novatchkova, K. Schernhuber, M. Fellner, A. Azaryan, M. Radolf, A. Stark, K. Keleman, B. J. Dickson, Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature* **464**, 287–291 (2010).
45. T. Chorna, G. Hasan, The genetics of calcium signaling in *Drosophila melanogaster*. *Biochim. Biophys. Acta* **1820**, 1269–1282 (2012).
46. P. J. Clyne, S. J. Certel, M. de Bruyne, L. Zaslavsky, W. A. Johnson, J. R. Carlson, The odor specificities of a subset of olfactory receptor neurons are governed by Acj6, a POU-domain transcription factor. *Neuron* **22**, 339–347 (1999).
47. T. Hummel, K. Krukkert, J. Roos, G. Davis, C. Klambt, *Drosophila* Futsch/22C10 is a MAP1B-like protein required for dendritic and axonal development. *Neuron* **26**, 357–370 (2000).
48. R. J. Kittel, C. Wichmann, T. M. Rasse, W. Fouquet, M. Schmidt, A. Schmid, D. A. Wagh, C. Pawlu, R. R. Kellner, K. I. Willig, S. W. Hell, E. Buchner, M. Heckmann, S. J. Sigrist, Bruchpilot promotes active zone assembly, Ca<sup>2+</sup> channel clustering, and vesicle release. *Science* **312**, 1051–1054 (2006).
49. J. E. C. Jepson, M. Shahidullah, D. Liu, S. J. le Marchand, S. Liu, M. N. Wu, I. B. Levitan, M. B. Dalva, K. Koh, Regulation of synaptic development and function by the *Drosophila* PDZ protein dyschronic. *Development* **141**, 4548–4557 (2014).
50. D. F. Eberl, R. W. Hardy, M. J. Kernan, Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. *J. Neurosci.* **20**, 5981–5988 (2000).

51. T. Kitamoto, X. Xie, C. F. Wu, P. M. Salva terra, Isolation and characterization of mutants for the vesicular acetylcholine transporter gene in *Drosophila melanogaster*. *J. Neurobiol.* **42**, 161–171 (2000).
52. A. E. Dubin, M. M. Liles, G. L. Harris, The K<sup>+</sup> channel gene ether a go-go is required for the transduction of a subset of odorants in adult *Drosophila melanogaster*. *J. Neurosci.* **18**, 5603–5613 (1998).
53. G. E. Stilwell, R. H. French-Constant, Transcriptional analysis of the *Drosophila* GABA receptor gene resistance to dieldrin. *J. Neurobiol.* **36**, 468–484 (1998).
54. A. Goto, T. Kadowaki, Y. Kitagawa, *Drosophila* hemolymph gene is expressed in embryonic and larval hemocytes and its knock down causes bleeding defects. *Dev. Biol.* **264**, 582–591 (2003).
55. M. Lindgren, R. Riazi, C. Lesch, C. Wilhelmsson, U. Theopold, M. S. Dushay, Fondue and transglutaminase in the *Drosophila* larval clot. *J. Insect Physiol.* **54**, 586–592 (2008).
56. M. Sugumar, Comparative biochemistry of eumelanogenesis and the protective roles of phenoloxidase and melanin in insects. *Pigment Cell Res.* **15**, 2–9 (2002).
57. N. Buchon, M. Poidevin, H. M. Kwon, A. Guillou, V. Sottas, B. L. Lee, B. Lemaître, A single modular serine protease integrates signals from pattern-recognition receptors upstream of the *Drosophila* Toll pathway. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12442–12447 (2009).
58. K. P. Rehorn, H. Thelen, A. M. Michelson, R. Reuter, A molecular aspect of hematopoiesis and endoderm development common to vertebrates and *Drosophila*. *Development* **122**, 4023–4031 (1996).
59. P. Irving, J. M. Ubeda, D. Doucet, L. Troxler, M. Lagueux, D. Zachary, J. A. Hoffmann, C. Hetru, M. Meister, New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cell. Microbiol.* **7**, 335–350 (2005).
60. C. G. A. Moreira, A. Jacinto, S. Prag, *Drosophila* integrin adhesion complexes are essential for hemocyte migration in vivo. *Biol. Open.* **2**, 795–801 (2013).
61. K. Nagaosa, R. Okada, S. Nonaka, K. Takeuchi, Y. Fujita, T. Miyasaka, J. Manaka, I. Ando, Y. Nakanishi, Integrin  $\beta$ v-mediated phagocytosis of apoptotic cells in *Drosophila* embryos. *J. Biol. Chem.* **286**, 25770–25777 (2011).
62. G. Campbell, H. Goring, T. Lin, E. Spana, S. Andersson, C. Q. Doe, A. Tomlinson, RK2, a glial-specific homeodomain protein required for embryonic nerve cord condensation and viability in *Drosophila*. *Development* **120**, 2957–2966 (1994).
63. W. C. Xiong, H. Okano, N. H. Patel, J. A. Blendy, C. Montell, repo encodes a glial-specific homeo domain protein required in the *Drosophila* nervous system. *Genes Dev.* **8**, 981–994 (1994).
64. F. Sieglitz, T. Matzat, Y. Yuva-Aydemir, H. Neuert, B. Altenhein, C. Klämbt, Antagonistic feedback loops involving Rau and Sprouty in the *Drosophila* eye control neuronal and glial differentiation. *Sci. Signal.* **6**, ra96 (2013).
65. Z.-Y. Ding, Y.-H. Wang, Z.-K. Luo, H.-F. Lee, J. Hwang, C.-T. Chien, M.-L. Huang, Glial cell adhesive molecule unzipped mediates axon guidance in *Drosophila*. *Dev. Dyn.* **240**, 122–134 (2011).
66. A. Volkenhoff, A. Weiler, M. Letzel, M. Stehling, C. Klämbt, S. Schirmeier, Glial glycolysis is essential for neuronal survival in *Drosophila*. *Cell Metab.* **22**, 437–447 (2015).
67. T. Awasaki, R. Tatsumi, K. Takahashi, K. Arai, Y. Nakanishi, R. Ueda, K. Ito, Essential role of the apoptotic cell engulfment genes draper and ced-6 in programmed axon pruning during *Drosophila* metamorphosis. *Neuron* **50**, 855–867 (2006).
68. J. M. MacDonald, M. G. Beach, E. Porpiglia, A. E. Sheehan, R. J. Watts, M. R. Freeman, The *Drosophila* cell corpse engulfment receptor draper mediates glial clearance of severed axons. *Neuron* **50**, 869–881 (2006).
69. S. W. Miller, T. Avidor-Reiss, A. Polyanovsky, J. W. Posakony, Complex interplay of three transcription factors in controlling the tormogen differentiation program of *Drosophila* mechanoreceptors. *Dev. Biol.* **329**, 386–399 (2009).
70. J. Kavalier, W. Fu, H. Duan, M. Noll, J. W. Posakony, An essential role for the *Drosophila* Pax2 homolog in the differentiation of adult sensory organs. *Development* **126**, 2261–2272 (1999).
71. Y. D. Chung, J. Zhu, Y. Han, M. J. Kernan, nompA encodes a PNS-specific, ZP domain protein required to connect mechanosensory dendrites to sensory structures. *Neuron* **29**, 415–428 (2001).
72. D. M. Duncan, E. A. Burgess, I. Duncan, Control of distal antennal identity and tarsal development in *Drosophila* by spineless-aristapedia, a homolog of the mammalian dioxin receptor. *Genes Dev.* **12**, 1290–1303 (1998).
73. A. Bitan, G. M. Guild, D. Bar-Dubin, U. Abdu, Asymmetric microtubule function is an essential requirement for polarized organization of the *Drosophila* bristle. *Mol. Cell. Biol.* **30**, 496–507 (2010).
74. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
75. J. Alquicira-Hernandez, A. Sathe, H. P. Ji, Q. Nguyen, J. E. Powell, scPred: Accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).

**Acknowledgments:** We thank F. Lamanna for advice on snRNAseq, M. Sémon for support on statistical analysis, C. Sinigaglia for sharing unpublished RNAseq data, F. lamunno and the electron microscopy facility of the Stazione Zoologica Anton Dohrn for support on SEM, B. Gillet and S. Hughes for support on next-generation sequencing, and the Centre Technologique des Microstructures of the Université Claude Bernard Lyon 1 for help in SEM sample preparation. We are grateful to E. Houliston and M. Akam for comments on the manuscript and to numerous colleagues for additional feedback. **Funding:** This work was supported by the Marie Curie ITN program EvoCell (H2020-MSCA-ITN-2017 #766053), the European Research Council grant ReLIVE (ERC-2015-AdG #694918), funded under the European Union Horizon 2020 program, and by a doctoral fellowship from Boehringer Ingelheim Fonds. **Author contributions:** A.A. performed the snRNAseq experiments. A.A. and M.P. analyzed the snRNAseq data. J.M.M. provided training and critical input on the snRNAseq work. Ç.Ç. performed the work on sensory organs, including the sensory function assay, and immunostainings. M.P. performed the statistical analyses. M.A. and M.P. conceived and supervised the study. All the authors contributed to writing and revising the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The R code and input files used for analyses and to produce figures are provided in <https://zenodo.org/record/692881> (DOI: 10.5281/zenodo.6566786).

Submitted 5 January 2022

Accepted 11 July 2022

Published 24 August 2022

10.1126/sciadv.abn9823