Transcriptional Characterization of Neuron Differentiation in *Hydra vulgaris*

By

ABBY SARAH PRIMACK

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry, Molecular, Cellular, and Developmental Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Celina Juliano, Chair

Bruce Draper

Anna La Torre

Jeff Farrell

Committee in Charge

2023

ACKNOWLEDGEMENTS

I cannot begin to express my thanks to my advisor Celina Juliano, who has been an incredible mentor, friend, and ally. Celina, you have supported me as a whole person through my best and worst times, and I don't have enough words to tell you how much that means to me. Thank you for creating a healthy, fun, and welcoming lab environment where everyone can thrive and be their best selves. To my lab mates, I am extremely grateful to have worked alongside all of you during our time together in the lab. Thank you especially to lab alumni Jack Cazet, Yash Abeykoon, Adrienne Cho, and Brandon Reyes-Chavez for making my time in lab a joy. It has been an absolute honor to work with all of you.

I would like to extend my sincerest gratitude to my committee members Jeff Farrell, Bruce Draper, and Anna La Torre. Thank you for guiding my research and for the hours and hours spent discussing UMAPs and trajectories and ec5 differentiation. Special thank you to Charlie David, Rob Steele, and Stefan Siebert for showing me the ways of the *Hydra*.

Enormous thanks to all the mentors who shaped me into the scientist I am today. To Jon Boyle, thank you for your infinite patience and for showing me how to stand up for myself. To my Earlham professors, especially Peter Blair, Wendy Tori, Amy Mulnix, and Bob Rosenberg, thank you for making science fun and for letting me explore at every level.

To my friends, near and far, I love you so much and would not be here today without you. Chicken Run Appreciation Society (Chiara Azzaretti, Sara Lepkoff, Maryanne Braine, and Stacey Hauff), you are the loves of my life, and I am so grateful that our paths crossed 13 years

ago. Taylor Heckman, thank you for all the time we spent together eating delicious meals and hanging upside down. To the Fellowship of the Nuts—Jasmin Green, Jasen Liu, Collin Gross, and Emmy—thank you for going on so many wonderful California adventures, cooking amazing house dinners, and for being my chosen family.

Special thanks to Lady Gaga's *Chromatica* album, my Spotify "writing playlist", and the eight-hour "Rainy Night Coffee Shop Ambience with Relaxing Jazz Music and Rain Sounds" YouTube track. From my first to my last writing assignment, you have been there every step of the way (on repeat).

Thank you especially to my family for raising me and for supporting me from near and afar. To my parents, Ann and Lazar, and my brother Seth, thank you for teaching me how to take care of myself and others, for supporting my love of science, and for letting me be me. I love you and would not be here without you, literally and figuratively.

To my beautiful wife and partner Clara. Thank you for the life we've built together; you are so special to me, and I could not have done this wild ride without you. I love you so much.

Finally, to all the *Hydra* I've met along the way. Thank you for helping me make a small dent in developmental biology.

ABSTRACT

The number of research organisms used to study nervous system development and function has substantially increased in recent years due to advances in technology, such as single cell RNA sequencing. This has enabled researchers to characterize the neuronal transcriptional diversity in a wide range of organisms. However, very few comprehensive transcriptional descriptions of adult nervous systems currently exist. Here, we provide a transcriptional analysis of the entire *Hydra vulgaris* adult nervous system. Although *Hydra* shares many of the same experimental advantages as well-established invertebrate models, such as small size, optical transparency, and ability to test gene function, it is also able to regenerate its entire nervous system from adult stem cells. This makes *Hydra* an excellent model for studying nervous system development and regeneration at the whole-organism level.

Although *Hydra* has been studied for over 300 years, the *Hydra* nervous system has never been fully defined on a molecular level. Prior to this work, the diversity, transcriptional profiles, and developmental lineages of *Hydra* neurons were unknown. In this dissertation, I present an extensive set of resources characterizing the homeostatic *Hydra* nervous system. **Chapter 1** describes the generation of a whole-animal single-cell RNA-seq atlas for *Hydra vulgaris*. This includes the first molecular map of the *Hydra* nervous system, identification of 11 transcriptionally distinct neuronal subtypes, and first validated molecular markers for endodermal neuronal subtypes. **Chapter 2** presents the most comprehensive transcriptional characterization of the adult *Hydra* nervous system to date. This includes both differentiation and transdifferentiation pathways and the identification of the chromatin states of *Hydra*'s 11

neuronal subtypes. The **appendix** includes additional manuscripts that were collaborative efforts stemming from my thesis work.

CONTENTS

INTRODUCTION

A major question in developmental neurobiology is how a fully functional nervous system capable of directing multiple behaviors forms *de novo*. To answer this question, it is important to investigate the process of 1) how neurons are made from stem cells and 2) how neurons work together to form neural circuits that control behaviors. Significant gaps in knowledge about this process include 1) how stem cells regulate the number and type of neurons they make and 2) how neurons properly integrate into neural circuits.

To investigate these processes and to characterize the development of an entire nervous system, it is advantageous to use an animal with a simple nervous system, such as is found in invertebrates. Although invertebrates have increasingly been used to study nervous system development, only the nematode *Caenorhabditis elegans* has a completely defined nervous system. This includes developmental lineages for all 302 of its neurons (Taylor et al. 2021, Reilly et al. 2020). However, although *C. elegans* offers a simple nervous system with optical clarity, it is unable to undergo adult neurogenesis and replace injured neurons. The freshwater cnidarian *Hydra vulgaris* is an excellent complementary model to *C. elegans* for understanding the mechanistic basis of nervous system development and regeneration. In addition to having a simple nervous system with optical clarity, *Hydra*'s nervous system has limitless regenerative capacity due to constitutively active stem cells. Furthermore, these stem cells continually replenish *Hydra*'s nervous system under homeostatic conditions, allowing for experimental access to neuron differentiation pathways in the steady state adult.

My dissertation takes advantage of this process to comprehensively describe the *Hydra* nervous system, including characterizing neuron diversity and molecular markers, building differentiation trajectories, and defining neural circuit participation. Before starting my work, researchers knew very little about the molecular complexity of the *Hydra* nervous system and how it was made. Previous work had only identified only a small number of genes involved in neural differentiation and regeneration. One major goal of my work was to comprehensively identify the suite of genes expressed during neurogenesis, determine their temporal ordering, and start the process of building gene regulatory networks to describe the differentiation of all neuron subtypes in *Hydra*.

The *Hydra* body plan is a radially symmetric hollow tube arranged around an oral-aboral axis. The mouth and tentacles (the "head") are located at the oral end, and the peduncle and basal disk (the "foot") are located at the aboral end. *Hydra* is formed by two epithelial monolayers, the ectoderm and the endoderm. *Hydra*'s nervous system consists of 3,000 – 5,000 neurons that make up two separate nerve nets, one embedded in each of the epithelial monolayers. These nets span the entirety of the body, with higher neuron densities at the oral and aboral ends. Neurons are part of the interstitial cell lineage and are derived from interstitial stem cells (ISCs) located in the ectoderm. Interstitial stem cells divide every ~24 hours, and ~10% of daughter cells undergo neuron differentiation via an intermediate progenitor state. Following injury, neuron progenitors migrate to the site of injury where they differentiate into neurons to repair the nervous system (Miljkovic-Licina et al., 2007).

In addition to differentiation from stem cells, neuron transdifferentiation has also been observed during nervous system maintenance. Transdifferentiation, or lineage reprogramming, is the occurrence of one terminally differentiated cell type changing into another terminally differentiated cell type. Due to tissue dynamics in *Hydra*, differentiated cells, including neurons, are continually displaced towards the oral and aboral ends. This led to the hypothesis that *Hydra* cells retain high developmental plasticity and can undergo transdifferentiation in response to changing developmental signals and axis positions (Bode, 1992). This has been well documented in *Hydra* gland cells, which are derived from stem cells as well as other gland cell populations. Siebert et al. (2008) characterized zymogen gland cells undergoing transdifferentiation into granular mucous gland cells as they move orally from the body to the head and concluded that "both stem cell-based mechanisms and transdifferentiation are involved in normal development and maintenance of cell type complexity in *Hydra*".

Although neuron transdifferentiation is hypothesized to occur in homeostatic animals due to tissue dynamics, this has never been shown as part of normal nerve net maintenance. Several studies performed on *Hydra* chemically depleted of ISCs have documented neuron transdifferentiation and found that regenerating neurons underwent changes in neuropeptide expression profiles and morphologies (Bode, 1992; Koizumi et al., 1988; Koizumi & Bode, 1986, 1991). One caveat of these studies is that *Hydra* lacking ISCs would be unable to make neurons *de novo* and may thus be forced to activate normally unused developmental pathways. My dissertation provides the first evidence of neuron transdifferentiation occurring in unmanipulated homeostatic animals.

In addition to understanding the developmental pathways active in *Hydra*, defining the number of neuron types present is essential to fully characterizing the nervous system. Prior to the work presented in this dissertation, the molecular complexity of the *Hydra* nervous system was unknown. Epp and Tardent (1978) identified the two nerve nets, however it was unclear how many subtypes participated in each nerve net and what their distribution was. Two studies broadly characterized ectodermal neurons into groupings via neuropeptide staining (Noro et al. 2019, Hansen et al. 2000), but markers for endodermal neurons were not identified. These groupings revealed that neuropeptide expression had discrete spatial boundaries, such as restriction of Hym176-C positive cells to the peduncle and Hym176-E positive cells to the tentacles. Neurons were further categorized into four non-overlapping neural circuits associated with specific behaviors, such as elongation and contraction, using live imaging of transgenic animals (Dupre and Yuste, 2017). Three of these circuits (RP1, CB, STN) reside in the ectoderm, and the fourth (RP2) resides in the endoderm. These circuit groupings, along with neuropeptide expression patterns, provide more clues into the complexity of the *Hydra* nervous system.

In this dissertation, I fill many of these knowledge gaps by providing a comprehensive transcriptional study of the adult *Hydra* nervous system. This includes identifying 11 neuron subtypes, including eight ectodermal and three endodermal populations, as well as three intermediate transdifferentiation populations. I build differentiation and transdifferentiation developmental trajectories and identify the chromatin landscape for each subtype. This work also begins to place neuron populations into neural circuits. The findings presented here provide foundational information for understanding the mechanisms that regulate nervous system regeneration and establish *Hydra* as a model for developmental neurobiology.

Chapter 1. Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution

This chapter was originally published in *Science*:

Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler, and Juliano CE. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science*. 2019; 365(6451):eaav9314. doi: 10.1126/science.aav9314.*

I made the following contributions to the work presented in chapter 1: I collected the two neuron-enriched single-cell RNA-seq libraries using FACS and validated the first markers of two endodermal neuron populations using promoter-driven transgenics. I also validated several endodermal epithelial and neuronal markers using RNA in situ hybridization, wrote figure captions, and provided feedback on the manuscript. I specifically contributed to figures 1.6, 1.8, 1.18, 1.42, 1.43, 1.44, 1.45, 1.46, and 1.47.

ABSTRACT

The adult *Hydra* polyp continually renews all of its cells using three separate stem cell populations, but the genetic pathways enabling this homeostatic tissue maintenance are not well understood. We sequenced 24,985 *Hydra* single-cell transcriptomes and identified the molecular signatures of a broad spectrum of cell states, from stem cells to terminally differentiated cells. We constructed differentiation trajectories for each cell lineage and identified gene modules and putative regulators expressed along these trajectories, thus creating a comprehensive molecular map of all developmental lineages in the adult animal. In addition, we built a gene expression map of the *Hydra* nervous system. Our work constitutes a resource for addressing questions regarding the evolution of metazoan developmental processes and nervous system function.

INTRODUCTION

Hydrozoans have been at the center of fundamental discoveries in developmental biology, including animal regeneration and early observations of stem cells (Trembley et al., 1744; Weismann, 1883). Among hydrozoans, the cell populations and lineage relationships are best characterized in the freshwater polyp *Hydra* (Figure 1.1 A-D) (Bode et al., 1987; Bosch and David, 1987; Campbell, 1979; David and Murphy, 1977; Sugiyama and Fujisawa, 1978). Homeostatic somatic maintenance of the adult *Hydra* polyp depends on the activity of the differentiation pathway for all cells, which are replaced approximately every 20 days (Richard D Campbell, 1967). *Hydra* has three cell lineages—endodermal epithelial, ectodermal epithelial, and interstitial—with each lineage supported by its own stem cell population (Figure 1.1 A-D) (Bosch et al., 2010). All epithelial cells in the body column are mitotic unipotent stem cells, resulting in continual displacement of cells toward the extremities. Epithelial stem cells differentiate to build the foot at the aboral end and the hypostome and tentacles at the oral end (Figure 1.1 A-C); differentiated cells are eventually shed from the extremities (Holstein et al., 1991). Multipotent interstitial stem cells (ISCs) give rise to the three somatic cell types of the interstitial lineage—nematocytes, neurons, and gland cells (Figure 1.1 D)—and can also replace germline stem cells (GSCs) if they are experimentally depleted (Bosch and David, 1987; David, 2012; Nishimiya-Fujisawa and Kobayashi, 2012) (Figure 1.1 D). The cnidarian-specific stinging cells, the nematocytes, can fire once and are then discarded; neurons and gland cells are closely associated with epithelial cells and thus are continually displaced and lost (Campbell, 1974). Interstitial cells are maintained by three mechanisms: (i) mitotic divisions of ISCs, progenitors, and gland cells (David, 2012); (ii) ISC differentiation into neurons, nematocytes, and gland cells (Bode et al., 1987; David and Murphy, 1977); and (iii) change in the expression and function of

neurons and gland cells with position (Koizumi and Bode, 1986; Siebert et al., 2008). Thus, cell identity in *Hydra* depends on coordinating stem cell differentiation and gene expression programs in a manner dependent on cell location. Understanding the molecular mechanisms that underlie cellular differentiation and patterning in *Hydra* would be greatly facilitated by the creation of a spatial and temporal map of gene expression.

We used single-cell RNA sequencing (scRNA-seq) to complement this extensive knowledge of *Hydra* developmental processes. We collected ~25,000 *Hydra* single-cell transcriptomes covering a wide range of differentiation states and built differentiation trajectories for each lineage. These trajectories allowed us to identify putative regulatory modules that drive cell state specification, find evidence for a shared progenitor state in the gland cell and neural differentiation pathways, and explore gene expression changes along the oral-aboral axis. Finally, we generated a molecular map of the nervous system with spatial resolution, which provides opportunities to study mechanisms of neural network plasticity and nervous system evolution. We have made the single-cell data available at the Broad Institute's Single Cell Portal. We anticipate that providing a comprehensive molecular map as a resource to the developmental biology and neuroscience communities will rapidly advance the ability of researchers to make discoveries using *Hydra*. Cnidarians such as *Hydra* hold an informative position on the phylogenetic tree as the sister group to bilaterians (Dunn et al., 2014) and largely have the same complement of gene families found in vertebrates (Erwin, 2009; Kortschak et al., 2003; Technau et al., 2005). Thus, this dataset, in combination with the existing cnidarian single-cell dataset for *Nematostella* (Sebé-Pedrós et al., 2018), offers the opportunity to identify conserved developmental mechanisms.

RESULTS

Single-cell RNA sequencing of whole *Hydra* reveals cell state transitions

Thirteen droplet-based single-cell RNA-seq (Drop-seq) libraries were prepared from dissociated whole adult *Hydra* polyps, and two neuron-enriched libraries were prepared using fluorescence-activated cell sorting (FACS)–enriched, green fluorescent protein (GFP)–positive neurons from transgenic *Hydra* (Figures 1.7 and 1.8 and Tables 1.1 and 1.2). We mapped sequencing reads to a reference transcriptome and filtered for cells with 300 to 7000 detected genes and 500 to 50,000 unique molecular identifiers (UMIs), resulting in a dataset with a detected median of 1936 genes and 5672 UMIs per cell (Table 1.3). We clustered the cells, annotated cluster identity using published gene expression patterns (Figures 1.1 E-F and 1.9), and further validated identities by performing RNA in situ hybridization experiments (Figure 1.10). In the clustering, cells separated according to cell lineage (Figure 1.1 E), and we observed the expected cell populations within each lineage (Figure 1.1 F). We captured cells in a wide range of differentiation states.

Several differentiation trajectories are evident even in the t-distributed stochastic neighbor embedding (t-SNE) representation, similar to findings in scRNA-seq studies performed in planarians (Fincher et al., 2018; Plass et al., 2018). For example, clusters that correspond to differentiated head and foot epithelial cells are connected to their respective body column stem cell clusters (Figure 1.1 F). Additionally, the interstitial stem cell clusters are connected to both neuronal and nematocyte progenitors (nematoblasts). We also identified distinct clusters for differentiated cells of the interstitial lineage—neurons, gland cells, nematocytes, and germ cells (Figure 1.1 F). We applied non-negative matrix factorization (NMF) to the full dataset and subsequently to all lineage subsets to identify modules of genes that are co-expressed within cell

populations (Figure 1.11) (Brunet et al., 2004; Farrell et al., 2018). As described below and in the supplementary materials, the recovered gene modules were used for doublet identification (see supplementary methods for discussion of doublet categories, Figures 1.8, 1.12-1.15), trajectory characterization, and the identification of transcription factor binding sites enriched in the cis-regulatory elements of co-regulated genes.

Trajectory reconstruction of epithelial cells reveals position-dependent gene expression

Epithelial cells constantly adjust their gene expression relative to their position as they divide in the body column and are displaced toward the extremities (Figure 1.1 A). To identify these position-dependent gene expression patterns, we performed trajectory analyses on subsets of endodermal and ectodermal epithelial cells (Figures 1.2 A-B and 1.16 A-C). We ordered cells along the oral-aboral axis by using the R package URD to generate branching trajectories for each lineage spanning from the foot (aboral) to the hypostome and tentacle (oral) as two separate endpoints (Farrell et al., 2018). URD connects cells with similar gene expression and uses simulated random walks to find gene expression trajectories between terminal cell populations and a starting progenitor cell population. This required removing biological and technical doublets from the epithelial cell subsets, which we accomplished by using NMF module co-expression to identify doublet signatures (see methods and Figure 1.13). To validate these differentiation trajectories, we visualized the spatial expression of several previously characterized genes and validated the expression of several uncharacterized genes by RNA in situ hybridization (Figures 1.2 C-M, 1.17-1.19).

We identified epithelial genes with variable expression along the oral-aboral axis, including differentially expressed gene modules identified by NMF (Figures 1.20-1.23). These spatially

and temporally resolved gene expression profiles for body column epithelial cells provide access to putative regulators of epithelial cell terminal differentiation at the oral and aboral ends, such as transcription factors and signaling molecules (Figures 1.2, 1.18, and 1.21). For example, we find differential expression along the body axis of previously uncharacterized genes in the Wnt, BMP (bone morphogenetic protein), and FGF (fibroblast growth factor) signaling pathways (Figure 1.2). Therefore, these data suggest candidate genes and pathways for functional testing to better understand oral-aboral patterning in *Hydra*.

Identification of multipotent interstitial stem cells and trajectory reconstruction of the interstitial lineage

We extracted 12,470 interstitial cells from the whole dataset (Figures 1.1 E and 1.16 D), performed subclustering, and annotated the clusters through the expression of known and new markers (Figures 1.3 A and 1.24). The t-SNE representation of interstitial cells showed evidence for ISC differentiation (Figures 1.3 A and 1.24 A-H). NMF analysis was used to identify gene modules associated with interstitial lineage differentiation pathways (Figure 1.25). We identified a population of cells that largely lack expression of differentiation gene modules (i.e., the putative multipotent ISCs) and used this cell population as the root in an URD trajectory reconstruction (Figure 1.25). *HvSoxC* (Hemmrich et al., 2012) was found to be expressed in transition states between candidate ISCs and differentiated neurons and nematoblasts, which suggests that the expression of this gene marks cells undergoing differentiation (Figures 1.3 B and 1.26). We attempted to identify transcripts specific to the putative ISC population and found only a single marker with no shared similarities to known proteins (Figures 1.3 C and 1.27 A). ISCs may therefore be largely defined by an absence of cell type–specific markers, similar to planarian cNeoblasts (Fincher et al., 2018). The URD reconstruction recovered a branching tree

of interstitial stem cell differentiation that resolves neurogenesis, nematogenesis, and gland cell differentiation (Figure 1.3 D). We performed double fluorescence in situ hybridization (FISH) to validate predicted transition states (Figures 1.3 E-F and 1.26-1.28).

The trajectory analysis of the interstitial lineage suggests that neuron and gland cell differentiation transit through a previously undescribed shared cell state (Figure 1.3 D), whereas nematogenesis is distinct. To test this result, we identified genes that are expressed in the progenitor state common to neural and gland cell differentiation, including *Myc3* (t18095) (Hobmayer et al., 2012) and *Myb* (t27424) (Figures 1.3 D-E and 1.27). We identified the spatial location of *Myb*-positive cells using FISH and found positive cells in both the endodermal and ectodermal layers (Figures 1.3 F and 1.28). A subset of *Myb*-positive cells co-express the neuronal marker *NDA-1* (Augustin et al., 2017), consistent with *Myb*-positive cells giving rise to neurons in both epithelial layers (Figures 1.3 E-F and 1.28). Furthermore, we found endodermal *Myb*-positive cells that co-express *COMA* (t2163), a gene expressed during gland cell differentiation and in all gland cell states (Figures 1.3 E-F, 1.27 E, and 1.28). ISCs reside in the ectodermal layer but are the source of both new gland cells and neurons in the endodermal layer (Bode, 1996; David, 2012). The data suggest the existence of two *Myb*-positive progenitor populations: one that stays in the ectoderm to give rise to neurons, and another that crosses the mesoglea to the endodermal layer and subsequently gives rise to endodermal neurons and gland cells (Figure 1.3 G). Finally, we find that many of the gene modules identified by NMF analysis were specific to each differentiation pathway with ordered expression in pseudotime (Figure 1.29), thus revealing gene expression changes that underlie differentiation in the interstitial lineage.

Subtrajectory analyses of interstitial cell types

We next explored the specification of different cell types within the interstitial lineage (Figure 1.1 D). First, we examined nematocytes, which contain one of the most complex eukaryotic organelles, nematocysts (Özbek, 2011); these are used to sting and immobilize prey. *Hydra* nematocytes each have one of four types of nematocysts: desmonemes, holotrichous or atrichous isorhizas, and stenoteles (Bode and Flick, 1976; Holstein, 1981). We identified one cluster of differentiated nematocytes, which contains nematocytes that harbor either desmonemes or stenoteles (Figure 1.3A, cluster "nematocyte," and Figure 1.30 A-F). In addition, we annotated the differentiation trajectories of nematoblasts and identified gene modules that are expressed as they produce these two types of nematocytes (Figures 1.3 A-D, 1.4 A, and 1.29 to 1.31). Although extensive work on nematocyst diversity has been facilitated by their extreme morphological and functional differentiation, little is known about nematocyte molecular diversity. The identification of genes that are differentially expressed between nematocytes harboring different nematocyst types (Figures 1.4 A and 1.29 to 1.31) provides a basis for understanding the specification and construction of these extraordinary organelles, which are the defining feature of Cnidaria.

Second, we analyzed gland cells, which are interspersed between endodermal epithelial cells. Gland cell numbers are maintained both by specification of new gland cells from ISCs and by mitotic divisions of differentiated gland cells (Bode et al., 1987). We were able to capture ISC differentiation into gland cells in the trajectory analysis (Figures 1.3 D and 1.32). Zymogen gland cells (ZMGs) are found throughout the body and transdifferentiate into granular mucous gland cells (gMGCs) when they are displaced into the head (Figure 1.4 B). Both of these cell types exhibit location-dependent changes in gene expression, and we captured these by building

linear trajectories along the oral-aboral axis (Figures 1.4 C, 1.20, and 1.33). We hypothesized that spumous mucous gland cells (sMGCs), a separate type of gland cells present in the head, may exhibit similar location-dependent gene expression profiles that were previously unappreciated. Indeed, reconstruction of a linear trajectory uncovered oral-aboral organization of gene expression in this cell type, including several oral organizer genes (such as *HyWnt1*, *HyWnt3*, *HyBra1*, and *HyBra2*) in the orally located sMGCs (Figures 1.4 D, 1.20, and 1.34). This raises the possibility that these cells participate in patterning the head. Overall, our analysis reveals a broad range of gland cell states in *Hydra* that can be achieved through multiple paths.

Finally, we explored the germ cell clusters recovered in the dataset. We excluded germline cells from the interstitial lineage tree reconstruction because differentiation of GSCs from ISCs does not typically occur in a homeostatic animal (Nishimiya-Fujisawa and Kobayashi, 2012); thus, we did not expect to observe transition states linking ISCs to GSCs. However, we did elucidate the spermatogenesis trajectory by analyzing the progression of cell states found in the two male germline clusters that were recovered in the subclustering of interstitial cells and used these data to identify and confirm several new male germline genes (Figures 1.20, 1.35, and 1.36). We identified two female germ cell clusters, which likely correspond to early and late female germ cell development (Figure 1.3 A). We performed in situ hybridizations for two genes (*HyFem-1*, *HyFem-2*) expressed in a subset of cells found in the early female germline cluster and found positive cells scattered throughout the body column, which we hypothesize are GSCs (Figure 1.4s E-H and 1.36 G-N). If so, this would be the first report of gene expression in *Hydra* that is

specific to GSCs and would allow for the study of GSCs in *Hydra* through the construction of GSC reporter transgenic lines.

Identification of putative transcriptional regulators of cell state–specific regulatory modules

The construction of differentiation trajectories allows us to determine the spatial and temporal expression patterns of transcription factors, and thus gain insight into the gene regulatory networks that control cell type specification. We aimed to identify the transcription factor binding sites shared by co-expressed genes and candidate transcription factors that may bind these sites. To identify co-expressed genes, we used NMF to interrogate a genome-mapped dataset and found 58 metagenes (i.e., sets of co-expressed genes) (Figures 1.37 and 1.38). To identify the putative regulatory regions of these co-expressed gene sets, we performed ATAC-seq (assay for transposase-accessible chromatin using sequencing) on whole *Hydra* (Figure 1.39). We identified regions of locally enriched ATAC-seq read density (peaks), which signify regions of open chromatin, and restricted the analysis to peaks within 5 kb upstream of the start codon of the genes in each NMF metagene (Figures 1.38 to 1.40). We then performed motif enrichment analysis to identify transcription factor binding sites that may control the expression of genes belonging to a metagene. We found at least one significantly enriched motif for each of 39 metagenes.

These metagenes had distinct sets of enriched motifs, suggesting differences in the transcription factor classes underlying various cell states (Figures 1.5 A and 1.41). For example, the paired box (Pax) motif is enriched in regulatory regions of genes expressed during early and mid-stages

of nematogenesis, the forkhead (Fox) motif is enriched at mid- and late stages, and the POU motif is enriched only in late stages. The B cell factor (EBF) motif is enriched in the female germline and the TCF motif is enriched in neurons and gland cells. Among epithelial cell states, motif enrichment is less tightly restricted to particular cell states. However, the ETS domain binding motif is enriched in metagenes expressed in endodermal and ectodermal epithelial cells in the extremities (tentacles and foot). Additionally, homeodomain (Otx and Arx) and bZip motifs are enriched throughout both epithelial lineages, and forkhead motifs appeared to be associated with genes expressed in endodermal epithelial cells (Figure 1.5 A). The enrichment of forkhead motifs in *Hydra* endoderm and *Nematostella* digestive filaments is consistent with a conserved function for forkhead transcription factors in cnidarian endodermal fate specification that is also found across bilaterians (Grapin-Botton and Constam, 2007; Sebé-Pedrós et al., 2018).

To determine the regulatory factors that may be coordinating gene co-expression modules, we identified transcription factors within each metagene that are predicted to interact with the binding site(s) enriched in that metagene using a combination of Pfam domain annotation and profile inference (JASPAR) (Figure 1.40). For 25 of the 39 metagenes with enriched binding motifs, we found one or multiple candidate transcription factors with putative function in cell fate specification (Table 1.5). For example, we found a metagene (wg32) that consists of 73 genes coexpressed during nematogenesis. A Pax transcription factor binding motif was significantly enriched in the potential regulatory regions near those genes, and the *Pax-A* transcription factor (t9974) is part of the metagene (Figures 1.5 A-B and 1.41). The results therefore strongly suggest that *Pax-A* functions during early nematogenesis. This is concordant

with a recent finding that *Pax-A* is required for nematogenesis during *Nematostella* development (Babonis and Martindale, 2017; Sebé-Pedrós et al., 2018). Similarly, we found evidence that an RX homeobox transcription factor (t22218) functions in basal disk development and an RFX transcription factor (t30134) functions in gland cell specification; the latter was also reported for *Nematostella* (Figure 1.5 C-D) (Sebé-Pedrós et al., 2018). Homeodomain transcription factor binding motifs are enriched in ectoderm tentacle genes (metagene wg71) and the analysis recovered aristaless-related homeobox gene *HyAlx* (t16456) as a regulator (Table 1.5 and Figure 1.17 A). A role for *HyAlx* during tentacle formation has been established previously (Smith et al., 2000). We provide all transcription factors that met our criteria as candidate regulators, including cases such as the basal disk where multiple TFs are both expressed in the proper context and predicted to bind an enriched motif (Table 1.5). Overall, we identified several candidates for regulators of *Hydra* cell fate specification.

A molecular map of the *Hydra* nervous system

The *Hydra* nervous system consists of two nerve nets, one embedded in the ectodermal epithelial layer and one embedded in the endodermal epithelial layer. Neurons are concentrated at the oral and aboral ends of the polyp (Bode et al., 1973). To determine the molecular nature of neuronal subtypes, we extracted neural progenitors and differentiated neurons from the dataset for subcluster analysis (Figures 1.16 E-F and 1.42). We identified 15 clusters: Three clusters consist of neuronal progenitor cells, expressing progenitor genes such as *Myb/Myc3*, and the remaining 12 clusters are differentiated neuronal subtypes (Figures 1.6 A and 1.43-1.46). To place these 12 neuronal subtypes into the ectodermal or endodermal nerve net, we performed TagSeq (Lohman et al., 2016) on separated tissue layers and conducted differential gene expression analysis to identify genes with significantly higher expression in the endodermal or ectodermal epithelial

layer (Figure 1.43, see methods). Because the neurons remained attached to the epithelia, differentially expressed genes included neuron-specific genes, which allowed us to score the neuronal clusters as ectodermal or endodermal (Figures 1.6 A and 1.43).

To determine the location of the ectodermal neuronal subtypes along the oral-aboral axis, we generated a list of neuronal markers and selected genes to test spatial location using a combination of new and previously published in situ expression patterns (Figures 1.6 A-B and 1.44-1.47). To test the endodermal identity of clusters en1 and en2, we examined *NDF1* (t14976, specific to cluster en1) and *Alpha-LTX-Lhe1a-like* (t33301, specific to cluster en2) expression by generating GFP reporter lines. For NDF1, GFP is expressed in endodermal ganglion neurons in the entire body except tentacles (Figures 1.6 C and 1.46 N-O). For *Alpha-LTX-Lhe1a-like*, GFP is expressed in sensory neurons along the body column in the endoderm (Figure 1.6 D-E). Therefore, the transgenic reporter lines confirm endodermal localization of clusters en1 and en2 and demonstrate our ability to identify specific biomarkers for each neuronal subtype. In summary, we have produced a molecular map of the *Hydra* nervous system that describes 12 molecularly distinct neuronal subtypes and their in situ locations.

DISCUSSION

We present an extensively validated gene expression map of *Hydra* cell states and differentiation trajectories, thus providing access to transcription factors expressed at key developmental decision points. Several recent studies have similarly demonstrated the value of conducting whole-animal (Fincher et al., 2018; Plass et al., 2018; Sebé-Pedrós et al., 2018) or whole-embryo scRNA-seq (Briggs et al., 2018; Farrell et al., 2018; Karaiskos et al., 2017; Sebé-Pedrós et al., 2018; Wagner et al., 2018) to uncover cell type diversity and the regulatory programs that drive

cell type specification. Conducting scRNA-seq on a diversity of organisms will provide insights into the core regulatory modules underlying cell type specification and the evolution of cellular diversity (Marioni and Arendt, 2017). Thus, our *Hydra* dataset provides an additional opportunity for comparisons to be made in an evolutionary context.

Analysis of *Hydra* by scRNA-seq uncovered new technical challenges, and we provide solutions to these challenges that will likely be applicable to many systems. For example, *Hydra* epithelial cells are highly phagocytic (Campbell, 1976), a phenomenon that has been observed in a variety of systems, and thus will likely present a challenge for interpretation of scRNA-seq results in future studies (Lu et al., 2011; Nakanishi and Shiratsuchi, 2004; Schafer et al., 2012). We implemented an approach that has been incorporated into URD, in which we use NMF as an unbiased method to identify anomalies in the data that likely represent cell doublets or phagocytic events. We envision that our approach could be applied to other systems and will be particularly useful in animals where existing expression data are limiting.

Our gene expression map for a dynamic and regenerative nervous system opens the door to understanding the molecular basis of neuronal plasticity and regeneration. Of the 12 neuronal subtypes we have identified, three (the endodermal neurons) were previously uncharacterized molecularly. Three distinct neuronal circuits have been described in *Hydra*: two in the ectoderm [rhythmic potential 1 (RP1) and contraction burst (CB)] and one in the endoderm [rhythmic potential 2 (RP2)] (Dupre and Yuste, 2017). These circuits are likely composed of ganglion neurons connected throughout the body. The characteristic localization of these circuits, combined with the in situ locations of the ganglion neuron molecular subtypes we identified (Figure 1.6 A), suggest the molecular identities of the neurons that constitute these distinct

circuits. We propose the following: (i) The endodermal neurons of cluster en1 (Figure 1.6 A-B) make up the RP2 circuit; (ii) the neurons of clusters ec3A, ec3B, and ec3C make up the RP1 circuit; and (iii) the neurons of clusters ec1A, ec1B, and ec5 make up the ectodermal CB circuit. This is supported by the observation that the RP1 circuit is active in the basal disk (cluster ec3A), whereas the CB circuit extends aborally only to the peduncle (cluster ec5) (Dupre and Yuste, 2017). Neuron subtype–specific transgenes, such as the two examples presented here, will provide powerful tools for experimental perturbations to test neuronal function and nervous system regeneration by enabling precise alterations to these neural circuits. Nervous system function in such engineered animals can be tested using newly developed microfluidic tools that allow for simultaneous electrical and optical recordings in behaving animals (Badhiwala et al., 2018).

The interstitial lineage differentiation trajectories provided several new insights. First, we identified a marker that may be specific to the multipotent stem cell population, which could provide a powerful tool for understanding stem cell function and fate decisions. Second, our data suggest the existence of a cell state that is shared by the neuron and gland cell trajectory (Figure 1.3 D). This interpretation is supported by the colocalization of neural and gland cell progenitors in several independent clustering analyses (Figures 1.1 F, 1.3 A, and 1.37) that consider different sets of variable genes and sets of cells, and by the overlap of gene modules for neurogenesis and gland cell differentiation (Figure 1.48). The shared stem cell of gland cells, neurons, and nematocytes suggests a shared evolutionary history of these cell types. The data further suggest that the evolution of nematocytes coincided with the emergence of a distinct progenitor. We thus propose a model in which multipotent ISCs first decide between a nematocyte or gland/neuron

fate and then a second decision is made by the common gland/neuron progenitor. This contrasts with previous models that posit a common neuron/nematocyte progenitor (Miljkovic-Licina et al., 2007). However, an alternative explanation is that gland and neuronal progenitors are separate populations that share early transcriptional events; thus, future fate-mapping experiments will be crucial. Additionally, our data suggest a model in which a bipotential gland/neuron progenitor born in the ectodermal layer, where multipotent ISCs reside, traverses the extracellular matrix to provide the endodermal layer with both gland cells and neurons (Figure 1.3 G).

Adult *Hydra* polyps, which are in a constant state of development, enable the capture of all states of cellular differentiation using scRNA-seq. An important future goal is to use scRNA-seq to rapidly assess the effect of mutations on all cell types (Farrell et al., 2018; Harland, 2018; Wagner et al., 2018). *Hydra* has a diversity of fate specifications from multiple stem cell types, yet is simple enough to be completely captured by a relatively small number of sequenced single cells from one life stage. Thus, we are now able to study organism-wide changes at a single-cell level in response to perturbations. The transcription factors that we identified at key developmental decision points are exciting candidates to test using this approach. In conclusion, this resource and the experimental approaches we describe open doors in multiple fields including developmental biology, evolutionary biology, and neurobiology.

ACKNOWLEDGEMENTS

We thank R. Monroy for collection of separated endoderm and ectoderm epithelial layers for TagSeq; R. Abhineet, J. Ho, Q. Li, N. Sierra, and A. Cho for wet lab validations by RNA in situ hybridization; Y. Wang for help with NMF analysis and helpful discussion; R. Steele, C. Dana,

and K. Glauber for kindly providing transgenic lines PT1, enGreen1, and nGreen; T. C. G. Bosch for kindly providing *Hydra vulgaris* AEP; L. Froenicke, V. Rashbrook, S. Ashtari, and E. Kumimoto of the UC Davis DNA Technologies Core for expert advice and assistance with sequencing; the MCB Light Microscopy Imaging Facility, which is a UC Davis Campus Core Research Facility, for the use of an Olympus FV1000 confocal microscope and M. Paddy for excellent support; C. David for providing numerous and plentiful insights on *Hydra* biology that helped with the interpretation of the data; B. Teefy for helpful discussion and moral support; and A. Schier, R. Steele, G. Wessel, B. Draper, S. Materna, J. Robinson, and C. Dunn for critical reading of the manuscript. Funding: The confocal microscope used in this study was purchased using NIH Shared Instrumentation Grant 1S10RR019266-01. This study was supported by DARPA contract HR0011-17-C-0026 (C.E.J.), UC Davis Start-up funding (C.E.J.), and NIH grant K99HD091291 (J.A.F.). Author contributions: S.S., J.A.F., J.F.C., and C.E.J. conceived the study; S.S., J.A.F., J.F.C., and C.E.J. wrote the paper with revisions by A.S.P., Y.A., and C.E.S.; S.S. and J.A.F. collected single-cell transcriptomes; Y.A., J.F.C., A.S.P., and S.S. performed ISH validation experiments; S.S. and Y.A. performed imaging; A.S.P. performed transgenic validation experiments of neurons and FACS sorting; J.F.C. performed ATAC-seq; S.S., J.A.F., and J.F.C. processed the raw data and conducted data analysis; S.S. and C.E.S. built transcriptome and gene models; J.A.F. built differentiation trajectories; J.F.C. and S.S. performed regulatory module analysis

MATERIALS AND METHODS

Code Availability

Analysis code is available in a git repository at *https://github.com/cejuliano/hydra single cell*.

Data availability

Raw Drop-seq reads, genes (rows) by cell (columns) data frames of unnormalized, unlogged transcripts detected per gene per cell for genome and transcriptome, raw RNA-seq data for epithelial specific libraries, a count matrix holding epithelial specific gene expression, raw ATAC- seq data and the consensus peak file have been submitted to the GEO repository; accession GSE121617. Single-cell data in processed and interactively browsable forms are archived in the Broad Single-Cell Portal (*https://portals.broadinstitute.org/single_cell/study/SCP260/stem-cell- differentiation-trajectories-in-hydra-resolved-at-single-cell-resolution*). Seurat objects for the whole transcriptome data set, the whole genome data set, the interstitial cell lineage subset, the epithelial cell lineage subsets, the neuronal cell subset, the URD tree object for the interstitial cell lineage, and URD spline objects for all trajectories that were reconstructed are also available for download on the Broad Single Cell Portal. Raw RNA-seq data used for de novo transcriptome assembly have been submitted to the Sequence Read Archive (SRA); Bioproject: PRJNA497966. The *Hydra* 2.0 genome assembly, the *Hydra* 2.0 gene models, and the de novo transcriptome for *Hydra vulgaris* AEP (aepLRv2) can be searched using Blast at the NIH *Hydra* 2.0 Project Portal (*https://research.nhgri.nih.gov/hydra/*). In addition, the Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GHHG00000000. The version described in this paper is the first version, GHHG01000000.

Aligned ATACseq reads for three replicates, the consensus peaks, and the *H. vulgaris* AEP transcripts can be displayed as tracks within the *Hydra* 2.0 Project Portal's genome browser.

Gene annotation and nomenclature

Published *Hydra* genes appear in italics in the text and figures. In the provided data objects (e.g Seurat objects, UMI count files), we added the mnemonic identifier of the best UniProt Knowledgebase match that was obtained by blastx searches of the Swiss-Prot database (E-value ≤ 1e-5) to the transcript IDs (e.g. t15331aep|FZD8_RAT). Throughout the manuscript, *Hydra* genes that have not been described are labeled in uppercase with the UniProt gene name followed by the transcript ID in parentheses. We use the transcript ID in case there is no *Hydra* identifier or Swiss-Prot ID. In a few instances, we found that sequences that did not return significant Swiss-Prot hits were annotated as part of NCBI's Annotation Process (*Hydra vulgaris* Annotation Release 102) and we used these annotations. Gene names and accession numbers for published *Hydra* genes that were used in this study are listed in Table 1.7. Annotations for uncharacterized genes are listed in Table 1.8. Gene ids that start with a leading "g" are gene model IDs from the *Hydra vulgaris* (strain 105) 2.0 genome assembly available at the *Hydra* 2.0 Project Portal.

Strains used in dissociations and spike-ins

Hydra were fed *Artemia* nauplii 2-3 times a week kept at 18℃. Animals were starved for 48h-96h prior to dissociation. Drop-seq libraries were created using non-budding or budding polyps of *Hydra vulgaris* AEP (Table 1.1), the strain that is used to generate transgenic *Hydra*. Whereas some *Hydra* strains reproduce exclusively asexually in the lab, *Hydra vulgaris* AEP frequently

undergoes sexual reproduction with the formation of eggs and sperm, providing the opportunity to sequence germline cells, and to obtain insights into germ cell development and differentiation. In total, 15 libraries were generated for *H. vulgaris* AEP, including one female specific (with developing egg patches) and one male specific (with testes) library. In selected libraries we spiked-in transgenic animals expressing GFP driven by an actin promoter predominantly in neuronal cells (nGreen, courtesy of Rob Steele), animals expressing GFP under the Cnnos-1 promoter (Hemmrich et al., 2012), or animals with developing eggs or sperm (see Table 1.1 for details).

Single cell isolation for Drop-seq

Hydra were dissociated using Pronase E (VWR, E629-1G) as described previously (Greber et al., 1992) at a concentration of ~75 units/mL. Optimal enzyme concentration was empirically determined for each batch of Pronase E. Dissociation was performed either in *Hydra* dissociation medium (Gierer et al., 1972; Greber et al., 1992) or in *Hydra* culture medium. 40-50 *Hydra* polyps were washed twice in sterile-filtered *Hydra* medium and transferred into a 1.5 mL tube. The medium was removed and 1 mL Pronase solution was added. Cells were dissociated for 90 minutes at room temperature (22-24°C) with gentle agitation on a nutator.

Cell Suspension Preparation

Following dissociation, cells were pipetted up and down 20 times using a pulled glass pipette or a 200 µl pipette to aid final tissue separation and cells were subsequently strained through a 70µm cell strainer (Corning #431751). Washing and pelleting of cells were conducted at room temperature as previously described (Greber et al., 1992). Initially, centrifugation speeds were

varied in an attempt to enrich for particular cell fractions. This caused slight compositional differences in the cell population across libraries, but enrichment for particular cell types was not achieved. The final spinning strategy included two spins at 300xg for 5 minutes at room temperature. The standard Drop-seq protocol uses PBS as cell buffer, a medium that would yield hypertonic conditions for *Hydra* cells (Online-Dropseq-Protocol-v.-3.1, *http://mccarrolllab.com/dropseq/*). Therefore, after each spin, cells were resuspended in either a salt-adjusted *Hydra*-PBS (10 mM PO4, 2mM NaCL, 0.1mM KCl, 0.05% BSA, pH7.6), in *Hydra*-PBS(iso) (10 mM PO4, 2mM NaCL, 2.7mM KCl, 0.05% BSA, pH7.6, iso-osmolar to *Hydra* dissociation medium), or in *Hydra* dissociation medium (*57*) containing 0.01% BSA (Table 1.1). Cells were strained a second time using a 70 μm or a 40 μm (Corning #431750) cell strainer prior to the determination of the cell concentration on a Fuchs- Rosenthal hemocytometer (Incyto).

Fluorescence Activated Cell Sorting (FACS)

Transgenic *Hydra vulgaris* AEP polyps expressing GFP predominantly in the cells undergoing neurogenesis and in mature neurons were used in FACS after dissociation into single cells as described above. The transgenic line (nGreen) was kindly provided by Rob Steele and was generated by microinjecting embryos with a plasmid containing the *Hydra* actin promoter driving GFP expression. A random integration event resulted in animals with a high fraction of neuronal progenitors and nerve cells positive for GFP protein (see below). Cells were sorted on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter, Miami, Fl, USA) with a 70 μm nozzle.

Drop-seq

Droplets were generated using a DSQ 3x9 array microfluidic part (Nanoshift) and a Drop-seq setup according to Macosko *et al*. 2015 (Online-Dropseq-Protocol-v.-3.1, *http://mccarrolllab.com/dropseq/*). Droplet sizes were determined for each new part that was used with fluorescent beads (P=S/2% 10uM Dragon Green, Bangs Laboratories) as described (*www.drop-seq.org*). Barcoded beads (Barcoded Beads SeqB; ChemGenes Corp., Wilmington, MA, USA, or Macosko sequence B – N13V, LGC Biosearch, Petaluma, USA) were prepared according to manufacturer's instructions. Compared to the ChemGenes beads, the LGC Biosearch beads had a 13 bp UMI as the only modification to the linker. Prior to the collection, cell syringes and tubing were blocked using *Hydra* PBS + 0.1% BSA. A magnetic mixing disc was inserted into the cell syringe to allow for manual cell mixing during the run and the cell pump was placed in a vertical position. Since cell suspension medium differed from the published protocol, salt was added to the lysis buffer to keep salt content in the droplet as close as possible to the original Drop-seq protocol (Table 1.1). Droplets were collected in 50 mL Falcon tubes and the target volume of combined aqueous flow varied between 0.8-1.2mL (1-1.2 ml cells and 1-1.2 ml of beads).

Droplets were broken immediately after collection and reverse transcription, exonuclease-treatment, and further processing were conducted as described previously (Evan Z. Macosko et al., 2015). For each library, three test PCRs (50 µl) containing bead equivalents of variable numbers of STAMPs (single-cell transcriptomes attached to microparticles) were conducted to determine the optimal cycle number for library amplification (Table 1.2). 35 µl of each test PCR were purified using Agencourt AMPure XP beads (Beckman Coulter, A63881, 21 µl beads

(0.6X) and 7 µl of H_2O for elution) and the DNA concentrations were determined using a Qubit 4 Fluorometer. A concentration between 400-1000 pg/µl was taken as optimal. PCR reactions were then conducted to amplify the remaining 1st strand cDNA with bead equivalents and optimal cycle numbers used in the test PCR (4 + 8-12 cycles, Table 1.2). 12 µl fractions of each PCR reaction were pooled, double-purified with 0.6X volumes of Agencourt AMPure XP beads (Beckman Coulter, A63881), and eluted in H_2O using 1/3rd of the bead volume. 1 µl samples from the amplified cDNA libraries were quantified with a Qubit 4 Fluorometer and library size distribution was verified on a BioAnalyzer High Sensitivity Chip (Agilent). 600 pg of library cDNA was fragmented and amplified (12 cycles) using the Nextera XT (Illumina) sample preparation kit and Drop-seq sequencing adapters. Libraries were double-purified with 0.6x volumes of AMPure XP Beads and quantified.

Bead and cell concentration

Beads and cells were loaded at concentrations specified in Table 1.2 and cells were counted using bright field microscopy. Cell concentrations were increased after observing low ratios (pilot libraries DS1_D1: 40.2%, DS1_P2: 41.9%) of recovered cells compared to anticipated STAMPs in pilot libraries. Whole *Hydra* cell suspensions display a wide range of cell sizes and include specialized cell types, and low recovery rates suggested that not all counted cells could be recovered as cells in a Drop-seq experiment. Forty-six percent of the cells in *Hydra* are nematoblasts (the cells that differentiate to become nematocytes) or nematocytes (Bode et al., 1973). Late nematocytes have a reduced cytoplasm and transcriptional activity when intact and cell membranes may be ruptured after nematocyst firing. We hypothesized that considering these cells in the countings may have decreased effective cell concentrations. Increasing cell

concentrations in selected libraries yielded improved recovery rates but some variability remained (Table 1.2). We also adjusted cell and bead concentrations according to the droplet size generated by the microfluidics part in use to achieve target concentrations. In case of the two FACS libraries of neurons, cell concentrations were based on the events counted in the sorting process.

Sequencing strategy

Nextera libraries were sequenced on Illumina Nextseq 500 sequencers using the NextSeq High Output v2 kit (75 cycles), using a custom primer and a custom paired end sequencing strategy (*49*) (read 1 20 bp, index read 8bp, read 2 60 bp in case of Chemgenes Macosko sequence B beads; read 1 25 bp, index read 8 bp, read 2 58 bp in case of Biosearch Macosko sequence B - N13V beads). Three to four Nextera libraries were pooled with a total of 8-12.8k anticipated STAMPs per NextSeq 500 run. Six NextSeq 500 flow cells were sequenced in total.

Preparation of a low redundancy de novo transcriptome for *Hydra vulgaris* (AEP)

Twenty *Hydra vulgaris* AEP polyps with an EF1-alpha promoter::GFP (enGreen1) transgene in endodermal epithelial cells (courtesy of Rob Steele and Catherine Dana) were cut to achieve tissue samples for four body compartments: tentacle, head, body and foot. Tissues from each compartment were pooled. A fifth tissue sample was prepared using 20 whole polyps. RNA was extracted using Trizol according to the manufacturer's instructions. Five sequencing libraries were prepared using TruSeq stranded mRNA (Illumina, RS-122-2201) and sequenced on half a lane of an Illumina HiSeq 2000 using a 100bp paired end sequencing strategy. Read 1 and Read 2 from the five libraries were concatenated and inserted into an agalma catalogue ((Dunn et al.,

2013), devel commit 0bf3d98, running BioLite commit 5302c8f). We determined the library insert size for use in the assembly process (command “agalma insert_size”, insert size: 270bp, standard dev: 91bp). Ribosomal reads were excluded running “agalma rrna” (read pairs examined: 129,125,848, read pairs kept: 128,996,009). Reads were then filtered for content (low complexity), Illumina adapters and quality using biolite script “bl-filter-illumina” using the default settings (pairs examined: 128,996,009, pairs kept: 81,148,392). This script excludes both reads in case one of the pairs fails in the filtering. We then corrected errors using the stand alone script ErrorCorrectReads.pl (ALLPATHS-LG release 48894 (Gnerre et al., 2011), pairs examined: 81,148,392, pairs kept: 76,372,001). Three stranded transcriptomes were assembled using Trinity 2.1.1 (command: Trinity --left readA.fastq --readB.fastq --seqType fq --group_pairs_distance 452 --min_contig_length 200 --SS_lib_type RF) (Grabherr et al., 2011). In the first two assemblies (assembly A, B) error corrected reads were assembled either with or without flag --min_kmer_cov 2. In the third assembly (assembly C) non-error corrected reads were assembled after trimming off the first 12bp of every read to remove potential sequence biases (Hansen et al., 2010) and using flag --min_kmer_cov 2. Assembly A included 83,512 Trinity transcripts, assembly B 92,474 Trinity transcripts and Assembly C 109,749 Trinity transcripts. Since we considered only uniquely mapped reads in downstream expression analyses the three assemblies were subsequently processed using script EvidentialGene tr2aacds.pl (v2017.12.21) to reduce redundancy ((Gilbert, 2013)). The final assembly (EvidentialGene okayset) comprised 38,749 sequences with an assembly N50 of 1.54 kb. Using BUSCO _v1.1b1.py (Simão et al., 2015) we found a reduced duplicate ratio of 15.3% (Complete Duplicated BUSCOs/Total BUSCO groups searched) compared to 46.6% in one of the starting transcriptomes. Since EvidentialGene introduces long transcript ids we simplified the naming

scheme to read tXXXaep numbering the transcripts in the reference from 1 through 38,749. Splice leader sequences (splice leader B: tatatacggaaaaaaacacatactgaaacttttagtccctgtgtaataag, (Stover and Steele, 2001)) were removed from the 5' end of each sequence using cutadapt 1.9.1 (Martin, 2011). 84 sequences were subsequently excluded from the deposited transcriptome because of sequence lengths shorter than 200bp. 24 additional sequences were excluded as they were flagged in the TSA contamination screen.

Reference annotation

The transcriptome was blasted against the Swiss-Prot database (E-value 1e-5, database date 03/08/2017) and Swiss-Prot IDs of best hits were added to the fasta header. We used the HMMER suite 3.1b2 (February 2015, *http://www.hmmer.org/)* and Pfam v31.0 database (Finn et al., 2016) to identify protein domains in the transcriptome and *Hydra* 2.0 genome gene models (*https://research.nhgri.nih.gov/hydra/*). Translated transcriptome sequences and the *Hydra* 2.0 genome gene models were scanned using the HMMER program hmmscan on default settings. For each domain model in Pfam, the domain was considered to be present in a protein sequence if it was detected with a HMMER independent expect-value equal to or below 1e-6. In the transcriptome assembly, we identified 435 transcripts with a DNA binding domain based on Pfam annotation, 424 of which are present in the scRNA-seq data set.

Bowtie2 reference for read alignment

The transcriptome assembly was screened for mitochondrial sequences using Blast and sequences were subsequently removed. The two *Hydra* mitochondrial chromosomes were added to the reference (accessions: NC_011220.1, NC_011221.1 (Voigt et al., 2008)). A sequence

dictionary and a .refFlat file were derived from the fasta file of sequences. The entire sequence from each transcript in the transcriptome reference was used as a gene model. Mitochondrial genes were annotated using the prefix MT_ following Voigt et al. (2008). A second reference was prepared for the *Hydra* 2.0 genome and gene models are available at *https://research.nhgri.nih.gov/hydra/*.

Read mapping, mapping stats, and generation of the expression matrix

We used Drop-seq tools v1.12 (*www.drop-seq.org*) to filter out barcodes with low quality bases, to trim poly(A) stretches and potential SMART adapter contaminants, and to add the cell and molecular barcodes to the sequences. The reads from the 15 libraries were then aligned to the de novo *Hydra vulgaris* AEP transcriptome and the *Hydra* 2.0 genome reference using Bowtie2 v2.2.6 (Langmead and Salzberg, 2012) with parameters --phred33 --very-sensitive -N 1 --reorder. The *Hydra* genome 2.0 (Chapman et al., 2010) was generated for a closely related strain *Hydra vulgaris* (strain 105), which resulted in reduced alignment rates. An average overall alignment rate of 78.4% was obtained when using the transcriptome reference compared to an average overall alignment rate of 60.5% for the genome. Drop-seq tools were further used to identify and correct bead synthesis errors, in particular, base missing cases in the cell barcode. Cell numbers were estimated by plotting the cumulative fraction of reads per cell against the cell barcodes and identifying the inflection point. The DigitalExpression script was used to obtain the digital gene expression matrix (DGE) for each sample. Reads not uniquely mapping to the reference were discarded by keeping the default READ_MQ of 10. Mapping to the genome allowed us to extract the regulatory sequences of co- expressed genes that were then used in motif enrichment analyses (see below).

Occurrence of biological multiplets and other doublet categories

Beside expected Drop-seq technical doublets (*i.e.* encapsulation of more than one cell together with a bead in a single droplet), we identified additional doublet categories that are due to tight physical association between cells that provide novel insight into *Hydra* biology. *Hydra* has a specialized cell that is localized in the ectoderm of the tentacles - the battery cell (Figure 1.1 D). This epithelial cell encloses multiple nematocytes and a neuron (Bode and Flick, 1976; Hobmayer and David, 1989; Hufnagel et al., 1985; Yu et al., 1985). The dissociation strategy did not allow for separation of the host cell and the integrated cells, and this type of association showed as a hybrid transcriptome in the data. Cells in clusters composed of battery cell multiplets are often positive for epithelial, nematocyte and neuronal markers, and are labeled as multiplet (mp) in the annotated t-SNE plot (Figure 1.1 D-F, clusters "battery cell", Figure 1.12).

In addition, mature nematocytes can be found mounted in ectodermal epithelial cells along the body column and neurons have also been found integrated within ectodermal cells outside the tentacles (Figure 1.1 B) (Bode and Flick, 1976; Dübel, 1989). We refer to a physical association of a neuron/nematocyte and an epithelial cell outside the tentacles that naturally occurs in homeostatic *Hydra* as an integration doublet (id) (Figure 1.1 F). We found unexpected co-localization of nematocyte gene expression with endodermal epithelial gene expression (e.g. Figure 1.13). With the exception of a single publication (Lyon et al., 1982), endodermal localization of nematocytes has not been established in *Hydra*. We attributed these doublet transcriptional signatures in part to phagocytic activity of epithelial cells. Phagocytic activity of epithelial cells in *Hydra* has been reported on multiple occasions (Campbell, 1976; Lyon et al.,

1982). Phagocytosis may occur across the mesoglea via phagocytic processes, even in homeostatic *Hydra*, as evidenced by cellular remains inside unbroken epithelial cells separated by maceration in wild type *Hydra (Campbell, 1976)*. Using transgenic lines expressing GFP in specific interstitial cell populations, we were able to find evidence in support of the hypothesis that neighboring cells are phagocytosed during the dissociation procedure (Figure 1.14). Phagocytic uptake occurs within 5 to 30 min post challenge and thus likely occurs before the animals are fully dissociated (McNeil, 1981).

Phagocytic uptake of cells or blebs has also been demonstrated in aggregation experiments (Seybold et al., 2016). This could potentially lead to spurious expression of marker genes from different lineages in ectodermal or endodermal epithelial cells. We interpreted co-expression of nematocyte and endodermal markers in part as a result of these processes. This suspected doublet type was labeled as a phagocytosis doublet (pd) in annotated t-SNE plots (Figure 1.1F). A clear distinction between the id (integration) and pd (phagocytosis) categories was in many cases not possible. Multiplets involving neurons could be demonstrated in FACS enriched cell populations (Figure 1.8. The line (nGreen) used for this experiment expresses GFP predominantly in the neuronal trajectory and mature neurons (Figure 1.49). Using FACS, we were able to collect both single neurons and larger GFP-positive cells that we determined to be multiplets composed of GFP-positive cells residing within epithelial cells (Figure 1.8. Ectodermal cells of the body column or endodermal cells are suggested as host cells in cases where co-integrated nematocytes are absent (see also Figure 1.14 G-H). The presence of multiple biological doublet categories as well as expected technical doublets (e.g. Drop-seq doublets) imposes analysis challenges. We therefore did not perform global doublet exclusion at an early stage, but rather evaluated

doublets at each stage of the downstream analyses and applied strategies at each stage to appropriately deal with these challenges. However, the discussed types of doublets/multiplets are often informative. Certain types of cells are frequently or exclusively found integrated and therefore the doublet/multiplet represents the only source of information. We found that clustering is often driven by genes expressed by the host epithelial cells. Marker genes of the host epithelial cells allow for spatial placement along the oral-aboral axis and thus can provide spatial information for genes expressed in the interstitial cell that was integrated or phagocytosed.

Gene/UMI filtering, doublet identification, and clustering of cells

We used the R package Seurat for cluster analyses and exploration of the data set (Satija et al., 2015). The complete analysis is available as a supplementary file (Supplementary analyses) and in the accompanying git repository as R markdown documents and knitted PDFs. To elucidate cell state specific gene and UMI metrics and to make an informed decision on suitable gene and UMI cut- offs, we performed an initial cluster analysis using permissive gene and UMI cut-offs (gene per cell cut-offs of > 200 and < 8k and UMI per cell cut-offs of > 400 and < 70k) (supplementary analysis SA01). We kept genes that were expressed in at least three cells and excluded cells that had more than 5% mitochondrial reads because this may indicate that a cell was stressed or dying. The data were normalized and scaled, and genes that varied more than expected for their expression level were identified. Principal components (PCs) were calculated on these variable genes which were then used in graph-based clustering followed by t-SNE dimensionality reduction. Throughout the analyses, we annotated cluster identity through the visualization of previously characterized genes and the use of publicly available RNA in situ

hybridizations results (Figures 1.9, 1.24, 1.44-1.45). Multiple markers were used to refine the annotations and additional RNA in situ hybridizations were performed to elucidate expression of biomarkers for selected clusters (e.g. Figures 1.10, 1.26, and 1.46). In the analyses, no clear gene and UMI cut-offs were suggested in elbow plots. We found that mature neurons and mature nematocytes presented the lowest gene and UMI numbers (Table 1.4) across libraries. Notably, Biosearch beads generated lower gene and UMI numbers per cell for comparable cell suspensions when compared to libraries with identical input that were generated using ChemGenes beads (supplementary analysis SA01). To keep neuronal cell states, we decided to increase the lower cut-offs to 300 genes and 500 UMIs, which retained all identified cell states represented as clusters. On the high end (>7000 genes) we found that cells contributed to multiple clusters, not suggesting that particular doublet categories could be excluded via cut-off selection. For the downstream analyses, we chose >300 < 7k genes, >500 UMI < 50k UMIs per cell as cut-offs except when noted otherwise.

In initial explorations after applying the final cut-offs, two clusters were identified with doublet expression signatures (supplementary analysis SA02). One category of cells showed expression for both endodermal epithelial cell and zymogen gland cell marker genes. Since zymogen gland cells reside in between endodermal epithelial cells, we interpreted this population as potential dissociation doublets. Gland cells are somatic differentiation products of the interstitial lineage (Bode et al., 1987), a lineage distinct from the epithelial cell lineages. The second category included cells positive for both endodermal and ectodermal epithelial cell markers. We interpreted these cells as potential Drop-seq doublets (two cells encapsulated with a single bead). Both categories of cells were excluded from downstream analyses. In both cases, phagocytic uptake may be an alternative source for the observed cross lineage co-expression of genes.

In the course of this study, two histone proteins were identified for which we could show male germline specific expression using RNA in situ hybridization (Figure 1.36 A-B). These histones were furthermore found to be expressed in cells of multiple clusters outside the germline clusters. 87.7% of cells that expressed either histone *H2BL1* (t11585aep) and/or *H10* (t38683) outside the male germline clusters originated from suspensions that intentionally had polyps with testes in them (Table 1.1, library 03-MA: male spike-in, 06-MA: all male library). We hypothesized that spermatids and sperm progenitors were not properly quantified when determining the concentration of the cell suspension, leading to higher than expected cell concentrations in the Drop-seq experiments and thus the generation of Drop-seq doublets. To ameliorate this, we excluded all cells outside the male germline clusters that showed histone expression (supplementary analysis SA02).

The resulting data set included 24,985 cells with a median of 1,938 genes and a median of 5,681 UMIs per cell. In case of all final cell sets, we calculated the JackStraw statistic and evaluated standard deviations of all PCs using function PCElbowPlot as implemented in Seurat (Satija et al., 2015). In addition, we performed parameter sweeps to determine the impact of PC cut-off selection on clustering and resulting differences related to batch effects. This led us to consider fewer PCs than suggested by the Seurat metrics in several instances by excluding lower ranking principal components. For the whole data set, we tested PCs from 1:19 through 1:37 with three different seeds (1 (default), 100, 4024) and three perplexities, 20, 30 (default), 40. The selected analysis and libraries for the final data set are summarized in tables S3 and S4. We performed NMF analysis (Figure 1.11 NMF analysis wt_K96, see below) on this set of cells to identify metagenes, sets of genes that are co-expressed.

Subclustering of epithelial cells

We subclustered and curated cells from all three cell lineages to obtain lineage specific t-SNE plots and in preparation for URD trajectory reconstruction. For the endodermal subclustering, epithelial cell clusters were extracted, with the exception of clusters composed of nematoblast/endodermal epithelial cell and nematocyte/endodermal epithelial cell doublets (pd and id clusters) (Figure 1.16, supplementary analysis SA03). To retain only the highest quality cells, we restricted the analysis to cells that expressed a minimum of 500 genes and 2k UMIs. NMF analysis was conducted to identify co-expressed genes in the endodermal epithelial cell subset (Figure 1.22, NMF analysis en_K40). We clustered the cells after regressing out the library specific (batch) effects as a source of variation by using the vars.to.regress argument in the Seurat function ScaleData(). Batch analysis revealed clusters composed of cells originating exclusively from a specific set of libraries (libraries 01-, 02- and 03-) and an endodermal metagene (en19) was found to be expressed in a batch specific manner. We excluded all cells with high scores for this metagene (accept.high=0.2) for downstream analyses. A set of endodermal epithelial cells received high scores for an endodermal metagene (en36) that included neuronal genes such as LWamide (*79*). The distribution of the cells was non-random, suggesting integrated neurons, phagocytic activity, or dissociation doublets as possible sources. We wanted to retain these cells with partial neuronal signatures for interrogation but did not want neuronal genes to play a role when clustering the cells. We therefore removed the neuronal genes of the metagene (en36) from the list of variable genes that were considered when calculating PCs prior to clustering, but excluded these cells prior to URD trajectory reconstruction since neuronal expression could interfere in the reconstruction of epithelial cell differentiation. 4,505 cells were retained for clustering considering PCs 1:14 (Figure 1.2A).

For the ectodermal subset, ectodermal epithelial cell clusters were extracted, with the exception of clusters composed of nematoblast/endodermal epithelial cell and nematocyte/endodermal epithelial cell doublets (pd and id clusters) (Figure 1.16, supplementary analysis SA03). NMF analysis was performed on this subset (ec_K76). To reduce the number of confounding effects, we considered cells from five selected libraries in the subclustering for ectodermal epithelial cells and cells with expression of more than 500 genes. The set of libraries included three libraries that were collected on two consecutive days using ChemGenes beads and *Hydra* culture medium in the dissociation (libraries 02-), the other two libraries considered were collected on the same day using Biosearch beads and dissociation medium in the dissociation (libraries 11-) (Table 1.1). This subset included 2,617 ectodermal cells and an additional NMF analysis was performed for this set of cells (Figure 1.23, NMF analysis - ec_K79). No obvious batch effects are observed when clustering cells from experiments 02- and 11- separately. To integrate cells from the two sets, we followed the approach described by Butler et al. (2018) using diagonal canonical correlation analysis (CCA) (*80*) considering canonical components 1:10 (Figure 1.2 B). Similar to what was observed in the endoderm, we found instances of neuron/epithelial cell doublets that did not interfere with biologically meaningful clustering of ectodermal cells (Figure 1.15). This suggested that the list of variable genes considered in the principal component analysis contained primarily epithelial genes.

Subclustering of cells from the interstitial lineage

Cells from the interstitial cell lineage were extracted from the full dataset and subjected to NMF analysis (Figures 1.16, 1.29, NMF analysis ic_K75). The full set of interstitial cells was analyzed without regression of library batches as a source of variability (supplementary analysis SA04). Metagenes from the whole dataset analysis were used to get insights into epithelial cell

expression signatures in cells of the interstitial subset. Two interstitial cell lineage clusters showed endodermal or ectodermal gene expression and were subsequently excluded from downstream analyses. PCs 1:31 were considered in the final clustering. Clusters were annotated using genes with known expression in cells of the interstitial lineage (Figure 1.24).

Subclustering of neuronal cells and neuron placement

To increase the overall number of neurons in our scRNA-seq data set, we used a transgenic line (nGreen) to collect GFP expressing neurons by fluorescence-activated cell sorting (FACS), generating two neuron-enriched Drop-seq libraries (Figures 1.8 and 1.49). We extracted neuronal progenitors and differentiated neuron clusters from the interstitial subset and performed a subcluster analysis (supplementary analysis SA05). FACS as an additional step in the workflow introduced library-specific effects (Figure 1.42). To integrate cells from FACS libraries and non-FACS libraries, we used canonical correlation analysis (CCA) (Figure 1.42) (Butler et al., 2018).

To get insights into the epithelial origin of each of the identified neuronal cell populations, we performed epithelium-specific TagSeq (see below) and identified genes that show enriched expression specific to epithelia. The tissue samples included epithelial cells of the respective epithelium and also all epithelium-associated interstitial cells. We obtained expression data for 13,995 transcriptome reference sequences, of which 3,055 genes were significantly differentially expressed in the endodermal epithelium and 2,859 genes were significantly differentially expressed in the ectodermal epithelium (adjusted p-value <0.05). We used the Seurat function AddModuleScore (Satija et al., 2015) to score each cell in the neuronal data set for sets of genes specific to the endodermal epithelium or to the ectodermal epithelium. This allowed us to place

three neuronal subtypes in the endodermal epithelium and nine neuronal subtypes in the ectodermal epithelium (Figure 1.43). We annotated neuronal clusters and determined their spatial location by the following methods: 1) interrogating RNA in situ patterns of known markers from the literature (Figures 1.44-1.45), 2), performing additional RNA in situ hybridization experiments (Figure 1.46), and 3) creating transgenic reporter lines using the regulatory regions of predicted biomarkers (Figures 1.6 C-E and 1.46 N-O).

Biomarkers

Positive biomarkers for clusters in the whole data set clustering, the interstitial lineage subclustering and the neuronal subclustering were identified using the Seurat function FindAllMarkers using min.pct=0.25 and the default parameters otherwise. Genes were filtered using an adjusted p-value cutoff of 0.01 and sorted by fold change. We report these markers in supplementary Table 1.9. For the neuronal subset we present a heatmap for the top 12 markers (Figures 1.6 B and 1.47).

Clustering of cells after mapping to the *Hydra* 2.0 genome

Drop-seq reads mapped to the *Hydra* 2.0 Genome assembly were processed using the *Hydra* 2.0 gene models. Analogous to the transcriptome filtering, we excluded cells expressing testes-specific histones that were not part of the germline cluster, as well as suspected cell doublets with co-expression for endodermal and zymogen gland cell or ectodermal markers, respectively. We performed graph-based clustering considering cells with gene cut-offs of >300 and <7k genes and UMI cut offs of >500 and < 50k (supplementary analysis SA06) (Figure 1.37). Principal components 1:30 were considered in the clustering. NMF analysis (wg_K84) was performed to identify co-expressed genes for downstream motif enrichment analyses and the

identification of key transcriptional regulators (see below). Genome and transcriptome mappings yielded comparable clustering results (Figure 1.1 F and Figure 1.37 B). Since the genome and the transcriptome data set contain the same cells we can visualize genome metagene cell scores on clusterings that were obtained using transcriptome data (Figure 1.5 B-D). In case of the male germline, we built differentiation trajectories using both our transcriptome and the genome data (Figure 1.35, see below). Interchangeable results underline the validity of both resources.

NMF analysis

Non-negative matrix factorization (NMF; (Brunet et al., 2004)) is a dimensionality reduction approach that treats an input data matrix (that does not contain any negative values) as the product of two smaller matrices. NMF algorithms attempt to find the two matrices that, when multiplied together, create the closest possible approximation of the original dataset. This approach generates a parts-based representation of the original data, with the underlying assumption that there are hidden variables (i.e. gene co-expression modules, or metagenes) that can explain most or all of the information found in the original dataset. More specifically for scRNA-seq data, NMF assumes that the gene expression profile of every cell can be completely explained as a linear combination of metagenes. Performing NMF on a single-cell expression matrix will result in one matrix that contains cell scores and another that contains gene scores. Cells with high scores express a metagene strongly. High gene scores reflect how well the expression of a gene corresponds to the expression of the metagene.

To identify metagenes in the single-cell expression data, we performed NMF using a previously published NMF framework (*https://github.com/YiqunW/NMF*) (Farrell et al., 2018). A sample analysis is provided as supplementary file in the git repository (supplementary analysis SA07).

The analysis was performed on log-normalized read count data for a set of variable genes using the run_nmf.py function with the following parameters: -rep 20 -scl “false” -miter 10000 -perm True - run_perm True -tol 1e-7 -a 2 -init “random” -analyze True. Each NMF analysis was repeated 20 times using different randomly initialized conditions, enabling us to evaluate reproducibility.

Because the optimal number of NMF metagenes that can describe a dataset cannot be determined *a priori*, this parameter (referred to as K) needs to be determined empirically. We initially performed each NMF analysis over a broad range of K values (typically from 10 to 120 by steps of 5). We then compiled the results from these separate runs using the integrate_and_output.sh script and determined the range of K values that yielded the highest number of informative metagenes (see details below). This then informed a second round of analyses using a narrower set of K values (ten total K values by steps of one), from which a final set of NMF analysis parameters were selected.

To guide the selection of an optimal K value, we identified the point at which increasing K no longer increased the number of meaningful metagenes. We considered a metagene to be meaningful if it reproducibly described the covariation of multiple genes. Using a K value at the point where the number of informative metagenes becomes saturated allows us to maximize the resolution at which we analyze gene co-expression.

Because metagenes that recur across replicates for a particular K will not necessarily be perfectly identical, we developed an approach to link metagenes from different replicates so that we could evaluate reproducibility on a metagene-by-metagene basis. The rationale for this approach was that reproducible and robust metagenes should reliably cluster together similar groups of cells in

independent replicate analyses. First, we clustered cells based on their highest scoring metagene (after normalizing all metagene cell scores to have a maximum value of 1). Then, metagenes were linked from one replicate to another by identifying the metagene in the second replicate that had the highest number of cells in common with the metagene in the first replicate.

To estimate a metagene's reproducibility, we created a metric (cluster reproducibility score) to evaluate the robustness of the metagene-based clustering. Specifically, we calculated the average proportion of cells that were clustered together in one replicate that were also clustered together in all other replicates. For example, a poorly reproducible metagene might group together 100 cells in one replicate, but in another replicate only cluster together 20 of the original 100 cells, giving the metagene a cluster reproducibility score of 0.2. In contrast, a highly reproducible metagene would have a score close to 1.

To estimate the approximate number of genes that drive a particular metagene, we determined the average number of genes whose score was within one order of magnitude of the top scoring gene across replicates. The criteria for an informative metagene were that it should have, on average, more than 10 genes, and should have a cluster reproducibility score above 0.6 (good metagene). Metagenes that did not fulfill these criteria were disregarded after final K selection (bad metagene). We performed NMF analysis on the full dataset mapped to both the transcriptome (wt, K=96) and the genome (wg, K=84). We also performed NMF on transcriptome- mapped interstitial (ic, K=75), ectodermal (ec, complete subset from all libraries - K=76, subset considered in subclustering - K=79), and endodermal (en, K=40) subclustered datasets. For each NMF analysis we provide cell scores, gene scores and the top top 30 genes for

each metagene in the git repository. Metagenes from all analyses (we present NMF ec76 for the ectoderm) can be visualized at the Broad Single-Cell Portal.

ATAC-seq

To generate *Hydra* ATAC-seq libraries, we made use of a modified version of the OMNI- ATAC protocol (Corces et al., 2017). For each biological replicate, five whole bud-free *Hydra vulgaris* (strain 105) polyps that had been starved for 2 days were transferred into 1 ml of chilled *Hydra* dissociation medium (Gierer et al., 1972) and homogenized with a glass dounce homogenizer using ~60 strokes. Cells were then transferred to a 1.5 ml centrifuge tube and spun at 4℃ for 5 min at 500g. The pellet was then resuspended in 50 µl of chilled resuspension buffer (RSB; 10 mM Tris-HCl - pH 7.4, 10 mM NaCl, 3 mM MgCl2) with 0.1% Tween-20, 0.1% NP-40, and 0.01% digitonin. Lysis was allowed to proceed on ice for 3 minutes and was subsequently halted by adding 1 ml RSB with 0.1% Tween- 20. To determine the concentration of nuclei in the suspension, 19 µl of the suspension was combined with 1 µl of 10 mg/ml Hoechst 33342 and loaded onto a Fuchs-Rosenthal hemocytometer. After determining the concentration, ~50,000 nuclei were transferred to a new 1.5 ml centrifuge and pelleted at 4℃ for 10 min at 500xg. The nuclear pellet was then resuspended in 50 µl of tagmentation solution (1x TD buffer, 33% PBS, 0.01% digitonin, 0.1% tween-20, 2.5 ul TDE1) and shaken at 37℃ for 30 minutes at 1000 rpm on a Eppendorf Thermoshaker C.

PCR cycle numbers for library amplification were determined as described (Buenrostro et al., 2015). 11 total PCR cycles were used to amplify the first and third replicates and 12 cycles were used for the second replicate. Libraries were size selected using SPRIselect beads to include

fragments between 100 and 700 bp. All three libraries were pooled at equimolar concentrations and sequenced on a single lane of an Illumina HiSeq4000 sequencer using 2 x 150 bp reads, resulting in a total of 171M read pairs.

Raw reads were filtered using trimmomatic (Bolger et al., 2014) with the following settings: -phred33 ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:2:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:32. This removed unmatched pairs, low quality reads, and adaptor sequences (3.5M reads were removed after trimming). The reads were then mapped to the *Hydra vulgaris* (strain 105) 2.0 genome using bowtie2 v2.2.6 (Langmead and Salzberg, 2012, p. 2) with the following settings: -X800 --very-sensitive-local --mm. Of the 168M filtered reads, 125M mapped to the genome (~74%). We identified mitochondrial reads by separately mapping filtered reads to the *Hydra vulgaris* (strain 105) mitochondrial genome using the same mapping parameters as above. The reads that mapped to the mitochondrial genome were then excluded from downstream analyses. Of the mapped reads, 35M were mitochondrial (~28%). Of the 90M mapped non-mitochondrial reads, 17M failed to pass the selected MAPQ threshold (≥ 3). We then identified and excluded PCR duplicates with the Picard MarkDuplicates function (*https://github.com/broadinstitute/picard*) using the default settings. Of the 73M mapped non-mitochondrial reads, 20M were marked as PCR duplicates, leaving a final total of 53M read pairs that were used for downstream analyses. Peaks were called with MACS2 (Zhang et al., 2008) using the following parameters: -g 9e8 -p 0.1 --nomodel. To determine which peaks were biologically reproducible, we made use of the irreproducible discovery rate (IDR) framework established by the ENCODE consortium (*https://github.com/nboley/idr*) using the following parameters: --rank p.value --soft-idr-threshold 0.1 --use-best-multisummit-IDR (Li et al., 2011).

For downstream analyses, we generated a consensus peak list consisting of all peaks that passed an IDR threshold of 0.1 for at least one pairwise comparison among the three biological replicates (Figure 1.39 A).

To assess the quality of the ATAC-seq dataset, we made use of several metrics proposed by the ENCODE consortium (*https://www.encodeproject.org/atac-seq/).* We evaluated reproducibility using both the rescue ratio and the self-consistency ratio. The rescue ratio is a measure of the difference between the number of peaks in the "conservative" (i.e. biologically reproducible) consensus peak set and the "optimal" consensus peak set. Ideal data sets should have a rescue ratio near 1, which indicates the biological replicates give results highly similar to what one would find with perfect replicates. To create the optimal peakset, pseudoreplicates were first generated by pooling reads from all biological replicates and randomly splitting them into three new read files. These pseudoreplicates were then used to find a consensus peak set using the same IDR methodology that was applied to the true biological replicates (see above). The optimal peak set had 77,358 peaks while the conservative peak set had 76,746 peaks (a rescue ratio of ~1.01) indicating good replication across the three samples.

The self-consistency ratio measures discrepancies between pairs of replicates, with a ratio of 1 representing a highly similar pair. To calculate the self-consistency ratio, self-pseudoreplicates are first generated within each replicate by randomly sorting the reads of each replicate into two new read files. The two self-pseudoreplicates are then used for an IDR analysis to generate an optimal peakset for each biological replicate. The self-consistency ratio is then determined by comparing the optimal peakset length between two biological replicates. This calculation is done for each pairwise comparison among the biological replicates and only the highest value is

reported. For our dataset, the self-consistency ratio was ~1.11 (replicate 1 optimal peakset: 60806 peaks, replicate 2 optimal peakset: 54978), indicating a highly consistent set of biological replicates.

Finally, to determine the signal-to-noise ratio for our data set, we calculated the transcription start site (TSS) enrichment score. Core promoters are depleted in nucleosomes, which is reflected in ATAC-seq data by an increase in read mapping density just upstream of the TSS. We calculated the TSS enrichment score by determining the average ratio between the read mapping density at the TSS and the read mapping density at the 100 bp that lie 1 kb both up- and downstream from the TSS for the 2000 most highly expressed genes in the Drop-seq data set. Biological replicates 1, 2, and 3 had scores of 7.9, 5.8, and 6.1 respectively, thus demonstrating a strong enrichment of mapped reads at the TSS (Figure 1.39 B).

Motif enrichment analysis

To identify potential regulators that coordinate transcription during homeostatic *Hydra* development, we used the following criteria: the candidate regulator should possess a DNA binding domain (DBD), its expression should be correlated with the expression of a suite of co-expressed genes, and the cis-regulatory regions of those co-expressed genes should be enriched in a motif that could plausibly be bound by the candidate regulator (Figure 1.40). Because the analysis required data on putative regulatory sequences for genes of interest, we used the genome- mapped version of the single cell dataset. Analysis code for the enrichment analysis is provided as a supplementary file (supplementary analysis SA08).

We identified putative transcription factors (TFs) in the *Hydra* 2.0 genome gene models by the presence of a Pfam DNA-binding domain (DBD). The list of considered Pfam DBDs was a modified version of a previously published set of Pfam domains, that was extended by adding selected domains (De Mendoza et al., 2013) (see supplementary analysis SA08). We identified 506 putative TFs in the genome gene models, for 386 of which we have expression data in the genome-mapped single cell dataset. We then used two different approaches to determine which DNA motifs could plausibly be bound by the predicted TFs. First, we used the JASPAR profile inference tool (Khan et al., 2018; Khan and Mathelier, 2018) to identify which of the potential DBDs in the putative transcription factors have significant (E- value < 1e-5) similarity to a DBD for which there is binding data in the JASPAR database—the rationale being that DBDs with similar protein sequences will have similar DNA binding preferences. This generated a list of 298 motifs linked to at least one gene model.

We noted that there were many instances when gene models with high confidence Pfam DBDs failed to yield any significant hits for a binding motif using the Profile Inference tool. Thus, we used a second approach with less stringent criteria to identify potential binding motifs for these additional TFs (Figure 1.40). Because each JASPAR motif is linked to a Swissprot entry that contains domain composition information, we were able to identify Pfam domains associated with the list of 298 motifs we initially generated. This allowed us to assign likely binding motifs to these additional putative transcription factors based on their Pfam annotations. Importantly, all motif-TF interactions predicted by JASPAR profile inference were also recovered using Pfam annotations; however, the Pfam-based approach also yielded additional potential transcriptional

regulators. Thus, results found using profile inference represent a conservative subset of the Pfam binding predictions.

To identify gene co-expression modules, we used the metagenes from the wg_K84 NMF analysis. We extended the NMF results to make use of all gene models for which there was expression data (the initial NMF was performed using a much shorter list of biologically variable genes). We therefore considered genes that were correlated (correlation score > 0.4) with the non-zero cell scores of a metagene of interest as members of that metagene.

We then identified putative upstream cis-regulatory sequences associated with the extended wg_K84 metagenes using the ATAC-seq dataset (see above). Biologically reproducible peaks that were within 5kb upstream of the transcription start site of genes belonging to the same extended metagene were grouped together for downstream motif enrichment analysis (Figure 1.39 C). We tested for the enrichment of 298 JASPAR motifs in the metagene peak lists using HOMER (Heinz et al., 2010) to identify motifs that might be distinctive of specific metagenes. We made use of the default settings in HOMER to generate control sequences (~50,000 percent GC and length-matched sequences pulled from random loci in the genome). Enrichment was calculated using a binomial distribution and a significance cut-off of 0.05 (adjusted p-value). Once we identified enriched motifs (Figure 1.5 A and Figure 1.41), we looked for putative regulators of the enriched motifs that had a correlation score > 0.3 for the metagene in question (Figure 1.40). TFs that were predicted to bind the enriched motifs and passed the correlation threshold were considered to be candidate regulators of metagene expression (Table 1.5). *Hydra* 2.0 gene models (g) were matched to corresponding transcripts (t) in the transcriptome reference

using Blast. Transcriptome IDs are presented in the main text and transcriptome data are visualized in Figure 1.5 B-D.

Trajectory analysis - URD

URD (Farrell et al., 2018) was used to uncover the transcriptional trajectories that cells traversed as they assumed their fates in the interstitial lineage and to analyze the spatial expression of genes in the ectodermal, endodermal, and gland lineages (supplementary analyses SA09-SA15).

Removal of outliers

For URD trajectory analyses, outliers that are poorly connected in the data often disrupt the diffusion map. Thus, we calculated a *k*-nearest neighbor network between cells, based on Euclidean distance calculated according to a non-restrictive set of variable genes (from the Seurat analysis) with 200 nearest neighbors. Cells were then removed based on either unusually high distance to their nearest neighbor (interstitial: >40, endoderm: >30, ectoderm: >26) or unusually high distance to their 20^{th} nearest neighbor, given their distance to their nearest neighbor (interstitial: $NN20 > 6 + 1.1 \times NN1$, endoderm: $NN20 > 3.5 + 1.1 \times NN1$, ectoderm: $NN20 > 3 + 1.1 \times NN1$).

Removal of doublets by NMF modules

In droplet-based single-cell RNA-seq techniques, such as Drop-seq, cells are loaded for encapsulation at a low concentration, such that most droplets contain only a single-cell. However, inevitably, a small proportion of those droplets will contain multiple cells (colloquially called “doublets”). Moreover, the proportion of doublets is unusually high in *Hydra* due to the phagocytic behavior of their endodermal and ectodermal lineages; this results in ‘biological

doublets' from encapsulation of a single cell that has engulfed a cell of a different type (see "Occurrence of biological multiplets and other doublet categories"). In trajectory reconstruction techniques, these can be particularly pernicious since they create spurious connections between distinct cell types through their transcriptional similarity to each of their constituent cell types. Decomposition of our expression data by non-negative matrix factorization (NMF) had produced distinct expression modules that characterized overarching cell states within our trajectories, so we reasoned that doublets could be identified and removed based on their expression of NMF modules based on two approaches (see "NMF analysis").

First, cells were removed from a given lineage if they strongly expressed modules of a *different* lineage. For both the endoderm and ectoderm lineages, the NMF decomposition that was performed on the full data aligned to the transcriptome ("wt_K96") was used and scaled such that each module's expression ranged from 0–1. The modules that were expressed in the non-endodermal or non-ectodermal cells were determined based on assignment of clusters to the endoderm or ectoderm using previously known markers. Cells were removed from the endodermal lineage if they expressed any non-endodermal module at a level > 0.125, and cells were removed from the ectodermal lineage if they expressed any scaled non-ectodermal module at a level > 0.125.

Second, cells were removed that strongly expressed multiple modules characteristic of different cell types *within* a lineage. Here, lineage-specific NMF decompositions were used: "en_K40" for the endoderm, "ec_K76" for the ectoderm, and "ic_K75" for the interstitial lineage. Modules in these decompositions that exhibited a strong cell-type signature were used ("good metagenes" as

defined in “NMF analysis” above), while those that exhibited a more general cell state or technical signature were not used (“bad metagenes”).

However, selecting the proper modules to use for eliminating doublets was critical. For instance, many modules encoded sequential cell states in a differentiation process; in this case, the modules would essentially be expressed in opposing gradients, such that along the differentiation process, cells would initially express module 1 strongly, then both modules weakly, then module 2 strongly. Module pairs of this nature should not be used to remove doublets, as they would incorrectly identify the transitions between cell states as doublets. Thus, to identify non-overlapping module pairs, modules were considered pairwise, and both the fraction of the data that expressed two modules above a high threshold (> 0.3) and the fraction of the data that expressed the same modules above a low threshold (> 0.15) were identified. If the two modules were expressed in an overlapping gradient (for instance, two modules that encoded sequential cell states in a differentiation process), then the fraction of cells that expressed both should increase dramatically as the threshold was lowered; if the two modules were expressed in a non-overlapping fashion, then the fraction of cells that expressed both would remain about the same. Thus, all pairs of modules were identified where the portion of cells that expressed both at a high threshold was <4.5% (interstitial cells), <7% (endoderm) or <8% (ectoderm) of the data, and the change in overlap when the threshold was lowered was <7% (interstitial cells) or <15% (endoderm, ectoderm). Then, all cells were removed that expressed both modules from such a pair at an intermediate level (> 0.25).

Determination of variable genes

For URD trajectory analyses, a more restrictive set of variable genes was calculated on each

subset of the data, as previously described (Farrell et al., 2018). Briefly, a curve was fit that related each gene's coefficient of variation to its mean expression level and represents the expected coefficient of variation resulting from technical noise, given a gene's mean expression value. Genes that exhibited a much higher coefficient of variation than expected for technical variability were used in downstream analysis to focus on genes that were likely to encode biological variability (endoderm: 1.25-fold higher; ectoderm: 1.4-fold higher; interstitial: 1.45-fold higher; granular mucous: 1.4-fold higher; spumous mucous: 1.35-fold higher; male germline: 1.4-fold higher).

In order to focus the trajectories on the relevant developmental processes in some tissues, major competing signals were identified and their genes were removed from the variable gene list. Since those genes were not used in downstream analysis, they then would not influence which cells were connected to each other, and would not affect the resultant trajectory structure. In the body column of the endoderm, a major stress response was observed that was encoded in two NMF modules ("wt2" and "wt92"). Thus, genes highly ranked in those modules were removed from the endoderm variable genes (top 60 genes from "wt2" and top 80 genes from "wt92", as determined by the elbow in a plot of genes' ranking within the module). In the interstitial lineage, a batch/stress effect that resulted from FACS enrichment of nGreen (actin::GFP) cells was observed. Canonical correlation analysis had been used to integrate the two batches (see "Subclustering of neuronal cells and neuron placement"), so genes that were highly ranked in the first six canonical correlation components were removed from the variable gene list. The absolute value of genes' loading was plotted, and the elbow in the curve was used to choose an approximate threshold (CC1: >27.5, CC2: >30, CC3: >25, CC4: >25, CC5: >20, CC6: >25).

Construction of interstitial lineage branching trajectory

We then used URD to reconstruct a branching trajectory tree in the interstitial lineage. Despite being part of the interstitial lineage, the germline was excluded from this analysis and processed separately (see "Construction of non-branching trajectories") because interstitial stem cells primarily give rise to germline stem cells only after damage, so we did not expect the relevant transitions to be present in the data. Additionally, fully differentiated nematocytes (cluster "nematocyte" in Figure 1.3 A) were excluded from the analysis because we could not unambiguously separate them into the four expected subtypes; instead, four separate populations of nematoblasts were identified and used as terminal fates. The interstitial lineage tree was constructed primarily as previously described (*24*), with differences described below.

The 'root' or starting point of the tree were the putative interstitial stem cells, chosen as three clusters from the data that largely lack expression of differentiation gene modules (Figure 1.25) and had low expression of *HvSoxC* (Swiss Prot SOX4), which was found as a general marker of interstitial differentiation. Terminal neural populations were chosen based on the clustering produced by Seurat on the entire interstitial lineage dataset (see "Subclustering of neuronal cells and neuron placement"), informed additionally by expression of nGreen (actin promoter::GFP), which labels neuronal populations. Terminal nematoblast populations were chosen from Infomap- Jaccard clusters produced by URD (with 40 nearest neighbors) based on (1) prior knowledge of nematogenesis, (2) their late pseudotime as assigned by URD, and (3) several differentially expressed genes, supporting the idea that there were four transcriptionally distinct populations. Terminal gland populations were chosen from the Infomap-Jaccard clusters produced by URD based on (1) prior knowledge of gland cell genesis, (2) prior knowledge of their spatial markers expressed at the oral and aboral ends of the animal, (3) their pseudotime

assignments, and (4) visitation of the rest of the gland cell data after performing biased random walks from different clusters.

A diffusion map was calculated using *destiny* (Haghverdi et al., 2016, 2015), using 100 nearest neighbors (approximately the square root of the number of cells in the data), and with a locally-defined sigma based on the distance to each cell's 5–7th nearest neighbors. The diffusion map was evaluated by plotting pairs of diffusion components to see that (1) it exhibited a clear structure that exhibited at least some major expected branching events, and (2) many terminal cell types could be found as spikes in at least one pair of diffusion components. Pseudotime was then computed using the simulated 'flood' procedure previously described (*24*), using the following parameters: *n* = 100*, minimum.cells.flooded* = 2. Biased random walks were then performed to determine the cells visited from each terminal population in the data as previously described (*24*), using the following parameters: *optimal.cells.forward* = 0*, max.cells.back* = 100*, n.per.tip* = 50000*, root.visits* = 1. The cells visited by random walks from each tip were visualized on the t-SNE projection, to ensure that the majority of the data was visited and that tips were chosen that were well connected to the data and followed a specific path through it. The branching tree was then constructed using URD's *buildTree*function with the following parameters: *divergence.method* = "preference"*, save.all.breakpoint.info* = TRUE*, cells.per.pseudotime.bin* = 25*, bins.per.pseudotime.window* = 10*, p.thresh* = 1e-3*, pref.thresh* = 0.5*, and min.cells.per.segment* = 10. This was similar to previously described (Farrell et al., 2018), with one notable exception—to determine when two distinct trajectories should fuse into a single branchpoint, the trajectories are divided into windows according to pseudotime, and their visitation frequency from walks started in each tip were compared; if the difference in visitation was not statistically significant, the two trajectories fused in that window. Previously,

the two-sample Kolmogorov-Smirnov (KS) test was used, but here a 'preference' test was used. Namely, each cell's preference for walks from the two considered trajectories was calculated as:

$$\text{Vistation}_1 - \text{Visitation}_2 \,/\, (\text{Visitation}_1 + \text{Visitation}_2)$$

Then, visitation was assumed to be significantly different if either: (a) the preference distribution was bimodal, as determined by Hartigan's dip test (Hartigan and Hartigan, 1985) or (b) the absolute value of the mean preference was > 0.5 (indicating that even if it is a unimodal distribution, the cells were primarily visited by walks from one tip).

Construction of endodermal and ectodermal simple branching trajectories

In order to find genes that are spatially regulated in the endoderm, a simple URD branching tree was constructed that began at the foot and ended in the head of the endoderm (in two separate populations—the tentacles and the hypostome). Clusters from a 30-nearest neighbor Infomap-Jaccard clustering were used as the root (a cluster in the foot) and tips (a cluster in the tentacle and a cluster in the hypostome). Two clusters of cells that were largely stress or contaminant clusters were removed. A diffusion map was calculated using *destiny (Haghverdi et al., 2016, 2015)* with 60 nearest neighbors (~ the square root of the number of cells in the data) and a global sigma of 6. Pseudotime was calculated using the 'flood' procedure previously described (Farrell et al., 2018) with 100 simulations, stopping when 2 or fewer cells were visited by a new simulation. Biased random walks (n = 50,000) were simulated with very permissive parameters of *optimal.cells.forward* = 0 and *max.cells.back* = 500, since the large number of cells represented a small number of distinct cell states. Then a branching tree was constructed from the two trajectories using URD's *buildTree* function with the following parameters:

divergence.method = "preference", *cells.per.pseudotime.bin* = 25, *bins.per.pseudotime.window* = 8, *p.thresh* = 0.001, and *min.cells.per.segment* = 10.

Similarly, in order to find genes that are spatially regulated in the ectoderm, a simple URD branching tree was constructed that began in the basal disk (the foot) and ended in the head of the ectoderm (in two separate populations—the tentacle and the hypostome). Clusters from a 30-nearest neighbor Infomap-Jaccard clustering were used as the root (a cluster in the basal disk) and tips (a cluster in the tentacle and a cluster in the hypostome). A diffusion map was calculated using 40 nearest neighbors (~ the square root of the number of cells in the data) and *destiny*'s local sigma approach (Haghverdi et al., 2016, 2015), where the sigma for each cell is determined by the distance to its nearest neighbors. Pseudotime was calculated using the 'flood' procedure with 100 simulations, stopping when 2 or fewer cells were visited by a new simulation. Biased random walks (n = 50,000) were simulated with standard parameters of *optimal.cells.forward* = 20 and *max.cells.back* = 40. Then a branching tree was constructed from the two trajectories using URD's *buildTree*function with the following parameters: *divergence.method* = "preference", *cells.per.pseudotime.bin* = 25, *bins.per.pseudotime.window* = 8, *p.thresh* = 0.05, and *min.cells.per.segment* = 10.

Construction of non-branching trajectories for gland cells and male germline

Gland cell numbers are maintained both by specification of new gland cells from ISCs and mitotic divisions of differentiated gland cells (Bode et al., 1987). One type of gland cells, zymogen gland cells (ZMGs), are found throughout the body, and exhibit location-dependent morphological and gene expression changes as they are displaced along the body column (Augustin et al., 2006; Guder et al., 2006), including transdifferentiation into granular mucous

gland cells (gMGCs) when they are displaced into the head (Siebert et al., 2008). Thus, in order to analyze spatial expression in the ZMG/gMGC population, URD was used to reconstruct a non-branching trajectory to represent cells' location along the oral–aboral axis (Figure 1.33). Another mucous gland cell population, the spumous mucous gland cells (sMGCs), are present primarily in the head of the animal; these cells do not exhibit significant morphological variation, and it was unknown whether they exhibited location-dependent gene expression changes. However, we noticed that some of these cells expressed genes that are traditionally associated with the oral organizer in *Hydra* (e.g. *HyWnt3*, *HyBra1)*, so URD was used to reconstruct a non-branching trajectory to explore spatial expression along the oral–aboral axis in this cell population (Figure 1.34).

In these unbranched trajectories, URD's *graphClustering* function was used to determine Infomap-Jaccard clusters with 20 nearest neighbors, from which a 'root' cluster was selected at the aboral end (glands) of the trajectory. Additionally, a cluster from the oral end was also selected to act as a root in a second pseudotime calculation (see below). A diffusion map was calculated using *destiny* with 75 nearest neighbors and global sigma of 6 (gMGC/ZMG), 5.86 (sMGC). The number of nearest neighbors was larger than standard due to the small number of cell states represented in each data set, and the sigma was determined in reference to *destiny*'s auto- detected global sigma parameter for each dataset. Pseudotime was then determined from the root cluster as previously described (*24*) using URD's *floodPseudotime* function with parameters: $n = 100$ and *minimum.cells.visited* = 2. For the granular mucous trajectory, pseudotime was calculated from both ends of the trajectory; both were then normalized to vary between 0 and 1, one was inverted, and then the two calculations were averaged. Cells were then analyzed based on their expression and pseudotime assignment.

We also analyzed differentiation in the male germline. This subset of cells was processed as a separate trajectory (since transitions from the interstitial stem cells to the male germline were not expected to be observed). Additionally, we reconstructed a trajectory using the same data that had been aligned to the genome, due to the concern that the transcriptome had been constructed primarily from *Hydra* that were not undergoing gametogenesis, so it was possible that germline-specific transcripts would not be well represented. However, both reconstructions yielded similar results (Figure 1.35). The unbranched trajectory was constructed as described for the gland cells, with pseudotime calculated starting from both the most undifferentiated and most differentiated

Finding genes that vary spatially or during differentiation

Depending on the trajectory, pseudotime is either a proxy for differentiation (male germline) or spatial location (endoderm, ectoderm, spumous mucous gland cells, granular mucous gland cells/zymogen trajectories). Genes were considered that were expressed (*i.e.* > 0) in at least 1% of each population. Cells were ordered according to pseudotime, split into groups of 5 cells, and the mean expression was determined for each group. A spline curve was fit to the mean expression vs. pseudotime relationship, using the *smooth.spline* function from R's *stats* package, with the parameter *spar* = 0.875 (ectoderm, granular mucous, spumous mucous, male germline genome), 0.8 (male germline transcriptome), or 0.9 (endoderm). Genes that vary in pseudotime (and thus either spatially or during differentiation) were then selected as those that: (1) were well fit by the spline curve (noise is usually poorly fit), with the sum of squared residuals per fit point < 0.2 (endoderm, ectoderm), < 0.045 (granular mucous/zymogen, spumous mucous, male germline transcriptome-aligned), or < 0.06 (male germline genome-aligned), (2) varied significantly in actual expression, with a spline curve that changes >0.5 (granular mucous/zymogen, spumous mucous, male germline) or > 0.75 (endoderm, ectoderm) in log2

expression value, (3) varied significantly in scaled expression, with a spline curve that varies 30–40% (granular mucous/zymogen, spumous mucous, male germline) or 33–48% (endoderm, ectoderm), requiring less variation for data that is better fit by the splines, and (4) were fit better by the spline curve than a straight line with slope 0, with the ratio of the sum of squared residuals > 1.19 (granular mucous/zymogen, spumous mucous) or > 1.14 (male germline). Additionally, the varying transcription factors in each tissue were identified as the intersection of the varying genes and a list of *Hydra* transcription factors (see "Motif Enrichment Analysis" for curation of this list).

Biomarker verification using colorimetric and fluorescent in situ hybridization

In situ hybridization was based on a previously published protocol (Grens et al., 1996). For each in situ, 30 *Hydra vulgaris AEP* polyps that had been starved for at least 2 days were relaxed in chilled 2% urethane in *Hydra* medium (HM) for 2-3 minutes and subsequently fixed overnight at 4℃ in 4% paraformaldehyde in HM. Fixative was removed with three 5-minute washes in PBS. The animals were then transferred to 1.5 mL tubes. All subsequent washes were performed in 1 mL volumes, at room temperature, on a rocker using gentle agitation unless otherwise indicated. To remove undesired pigmentation, animals were transferred to 100% MeOH via 5-minute washes first in 33% MeOH in PBS then 66% MeOH in water. Samples were then bleached for a minimum of 1 hour in 100% MeOH.

Samples were rehydrated with 66% MeOH in water for 5 min and 33% MeOH in PBS for 5 min followed by three 10-minute PBT washes. The samples were then digested for 6-10 minutes with 10 µg/mL proteinase K in PBT. The digestion was stopped with a quick wash in 4mg/mL glycine in PBT followed by a 10-minute wash in fresh glycine solution. Residual glycine was removed

with three 10-minute PBT washes. The samples were then washed twice with 0.1 M triethanolamine in PBT for 5-10 minutes, once with 3 µL/mL acetic anhydride in 0.1 M triethanolamine for 5 minutes, and once with 6 µL/mL acetic anhydride in 0.1 M triethanolamine for 5 minutes. This was followed by three 5-minute PBT washes. The samples were subsequently refixed with 4% paraformaldehyde in PBT for 1 hour. The fixative was removed by three 5-10 minute washes with PBT followed by two 5-10 minute washes with 2xSSC (300 mM NaCl, 30 mM sodium citrate). Preceding hybridization, samples were washed with 50% 2xSSC/50% Hybridization Solution (HS) [50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 1x Denhardt's solution, 100 µg/mL heparin, 0.1% Tween 20, and 0.1% Chaps] for 10 min, starting at room temperature then transitioning to the hybridization temperature of 56℃. The remaining pre-hybridization and hybridization steps were all carried out at 56℃. Samples were washed with HS for 10 min and then a prehybridization step was performed in HS with 10 µL/mL sheared salmon sperm DNA for 2 hours. For hybridization, digoxygenin-labeled probe was added to a final concentration of 3 ng/µl in HS with 10 µL/mL sheared salmon sperm DNA. Prior to dilution, the probe was denatured at 85℃ in a modified HS (50% formamide and 5x SSC) for 5-10 minutes. Hybridization of the samples occurred for ~60 hours at 56℃ with no agitation.

Following hybridization, unhybridized probe was washed out at 56℃ in a series of washes using HS, 75% HS–25% 2x SSC (300 mM NaCl, 30 mM sodium citrate), 50% HS–50% 2x SSC, and 25% HS–75% 2x SSC for 10 min each. This was followed by two 30-minute washes in 2xSSC with 0.1% Chaps, with the second wash starting at 56℃ and transitioning back to room temperature. This was followed by four 10-minute washes with MABT (100 mM maleic acid, 150 mM NaCl, 0.1% Tween 20, pH 7.5). Samples were incubated in MABT with 1% BSA for 1

hour at room temperature. Following this, samples were blocked at 4℃ for 2 hours in 500-750 µL blocking solution (80% MABT with 1%BSA and 20% sheep serum (Gemini Bio Products, #100-117). Finally, the samples were incubated with 500µL of a 1:2000 dilution of the appropriate antibodies (anti-DIG-AP for colorimetric and anti-DIG-POD/antiFITC-POD for fluorescent) in blocking solution at 4℃ overnight.

For colorimetric in situs, unbound antibodies were removed by two washes in MABT with 1% BSA and six MABT washes at room temperature for 20 min each. Samples were rinsed with NTMT (100 mM NaCl, 100 mM Tris-pH 9.5, 50 mM MgCl2, 0.1% Tween-20) for 5 minutes and transferred to 6 well plates. The NTMT was then refreshed with 20 ul/ml of nitro blue tetrazolium (NBT)/ 5- bromo-4-chloro-3-indolyl phosphate (BCIP) solution added. Staining proceeded for an empirically determined period of time and was subsequently stopped with three rapid PBT washes. In preparation for imaging, samples were transferred to 100% EtOH via 5-minute washes in 33% EtOH in PBT and then 66% EtOH in H2O. Samples were then incubated in 100% EtOH until the precipitate appeared blue. Samples were then rehydrated for 5 minutes in 66% EtOH in H2O, then in 33% EtOH in PBS, and then transferred to PBS. Samples were mounted in 80% glycerol for documentation.

For fluorescent ISH, unbound antibodies were removed with eight 20-minute washes in MABT with 1% BSA. Samples were then rinsed 2 times with 100 mM borate buffer (1:1 of 200 mM borate stock (pH 8.5): boric acid 200 mM, sodium chloride 75 mM, sodium tetraborate (borax) 25 mM) for 5 minutes each. The samples were stained with a tyramide solution (100 mM Borate Buffer, 2% dextran sulfate, 0.1% Tween-20, 0.003% H_2O_2, 0.15 mg/mL 4-iodophenol in DMSO, 1:100 dilution of Alexa Fluor 488 or 594 tyramide reagent) for 25 min and the reaction

was stopped by four rapid PBT rinses. To inactivate the peroxidase, samples were incubated in 100 mM glycine (pH 2.0) for 10 minutes, followed by five 5-minute PBT washes. For double-labelled in situs the blocking, antibody incubation, and tyramide reaction steps were repeated using reagents appropriate for the second probe. In preparation for imaging, samples were stained in 1:1000 DAPI in PBT for 30 min. Samples were then dehydrated through a gradient of 30%, 50%, and 80% glycerol in PBT each lasting at least 1 hour. Animals were mounted in 80% glycerol with 40 mM NaHCO3.

Immunohistochemistry

Hydra polyps incubated on ice for 20 minutes, and then relaxed in cold 2% urethane (Sigma) in *Hydra* culture medium. Animals were then fixed for 1 hour in 4% PFA in *Hydra* medium at room temperature. Following fixation, animals were washed 3x for 10 minutes in 1 mL PBS while gently rocking, and then permeabilized with 1 mL 0.5% Triton X-100 in PBS for 15 minutes. Animals were incubated in blocking buffer for 1 hour (1% BSA, 10% goat serum, 0.1% Triton X-100) and then incubated overnight at 4℃ in primary antibody (1:500 mouse anti-GFP , Roche #11814460001). Following three 10-minute washes in PBS/0.5% Tween/1% BSA, animals were incubated in secondary antibody (1:1000 Alexa Fluor 488 goat anti-mouse IgG in blocking buffer, Invitrogen A11001) and Phalloidin (1:200 Thermo Fisher Scientific) in the dark for 1 hour at room temperature. Animals were then washed 3x for 10 minutes while gently rocking in PBS. During the second wash, nuclei were stained using Hoechst (1:1000, Thermo Fisher Scientific). Slides were mounted using Prolong Diamond Antifade Mountant (Invitrogen)

Tissue layer separation and sequencing for differential gene expression analysis

Viable *Hydra vulgaris AEP* endoderm and ectoderm layers were obtained as described previously (Lesh-Laurie, 1983). *Hydra* (expressing RFP in endodermal epithelial cells and GFP in ectodermal epithelial cells (Glauber et al., 2013)) were allowed to relax for approximately 30 sec to 1 min in M solution (Muscatine, 1961) that was adjusted to pH 2.5. The head was then excised by a cut directly below the tentacle ring, followed by the excision of the lower peduncle. The remaining cylinder of tissue was transferred to Haynes solution (Davis et al., 1966). Shortly after being placed in Haynes solution, the ectoderm contracts until it forms a small ring surrounding a column of extended endodermal tissue (Figure 1.43). The two tissue layers were then separated using forceps, and each layer was immediately transferred to Trizol (Invitrogen) and stored at -80℃. Eight to ten tissue samples were pooled prior to total RNA extraction. Three separate endodermal and three separate ectodermal pools were processed. RNA was extracted according to the manufacturer's protocol and was subsequently treated with DNAse I and purified by extraction with phenol/chloroform. RNA yield per sample ranged from 0.8-1.6 μg. Total RNA was analyzed on a Bioanalyzer 2100 (Agilent). Slightly lower RIN numbers were obtained for endodermal RNA samples (7.1, 7.9, 7.3) compared to ectoderm (8.8, 8.5, 9.1). Sequencing libraries were prepped for TagSeq (Lohman et al., 2016) at the UC Davis DNA Technologies Sequencing Core using the QuantSeq 3‘ mRNA-Seq Library Prep Kit for Illumina (FWD) (Lexogen). Libraries were sequenced on an Illumina HiSeq 4000 using a single end 50 bp sequencing strategy. To process raw reads, Poly-A stretches and sequencing adapters were trimmed using the script bbduk.sh (Joint Genome Institute) using the following parameters: k=13 ktrim=r forcetrimleft=11 useshortkmers=t mink=5 qtrim=t trimq=10 minlength=20. Ribosomal reads were excluded using bbduk.sh and a set of *Hydra* rRNA sequences.

Epithelium specific gene expression

Epithelium specific expression levels were estimated using RSEM v1.2.31 (Li and Dewey, 2011) and bowtie v1.1.2 (Langmead et al., 2009) using our de novo transcriptome reference. Rsem-calculate-expression was run using option --forward-prob=1.0 for a strand-specific sequencing protocol. Average read numbers with at least one reported alignment were 70.3% for ectodermal libraries and 56.7% for endodermal libraries. The QuantSeq 3‘ mRNA-Seq Library Prep protocol generates a read close to the 3’ end of polyadenylated RNA and is therefore sensitive to 3’ incompleteness. RSEM isoform counts from all endoderm and ectoderm replicates were combined in a genes by treatment count matrix. Differential gene expression analysis was conducted using edgeR v. 3.20.9 (Robinson et al., 2010).

Generation of transgenic lines

The generation of transgenic *Hydra* lines was performed as previously described (Juliano et al., 2014; Wittlieb et al., 2006). Genome gene models were identified for transcripts t33301 and t14976 that were predicted to be expressed in distinct endodermal neuron subtypes. Two plasmids were generated for *Hydra* 2.0 gene model g15727.t1 (t33301) (NCBI(nr): PREDICTED: alpha-latrotoxin-Lhe1a-like, partial [*Hydra vulgaris*]) and g26087.t1 (t14976) (similarity to Uniprot Neurogenic differentiation factor 1, *Danio rerio*). 1,673bp (g15727.t1) or 2,039bp (g26087.t1) upstream ATG (*Hydra* 2.0 genome assembly) were used as putative promoters. Promoters were cloned in expression vector pHyVec13 (Addgene plasmid: #34796) using restriction sites BamHI and PstI. The resulting constructs had the neuronal promoter upstream of GFP with the extra actin amino acids at the amino terminus of GFP. Plasmids were prepared by Maxiprep (Qiagen, Valencia, CA) and eluted in RNase-free water. Plasmid DNA

was injected into embryos at a final concentration of 1 mg/mL using an Eppendorf FemtoJet 4x and Eppendorf InjectMan NI 2 microinjector (Eppendorf; Hamburg, Germany) under a Leica M165 C scope (Leica Microscopes, Inc; Buffalo Grove, Il).

Imaging

An Olympus FV1000 confocal microscope was used to document tissue following fluorescent RNA in situ hybridization and fluorescent immunohistochemistry. Images were processed in ImageJ (vers. 2.0.0-rc-69/1.52i) (Schindelin et al., 2012). Colorimetric in situ RNA hybridizations were documented using a Leica DM5000B microscope (camera Leica DFC310FX) or Leica M165C digital stereo microscope (camera MC170HD). Brightness was adjusted using Adobe Lightroom Classic CC (release 8.1). In case of Figures 1.4 F and 1.36 H-L, multiple shots along the *Hydra* body column were acquired and aligned using auto-align layers in Adobe Photoshop CC (release 20.0.1).

FIGURES

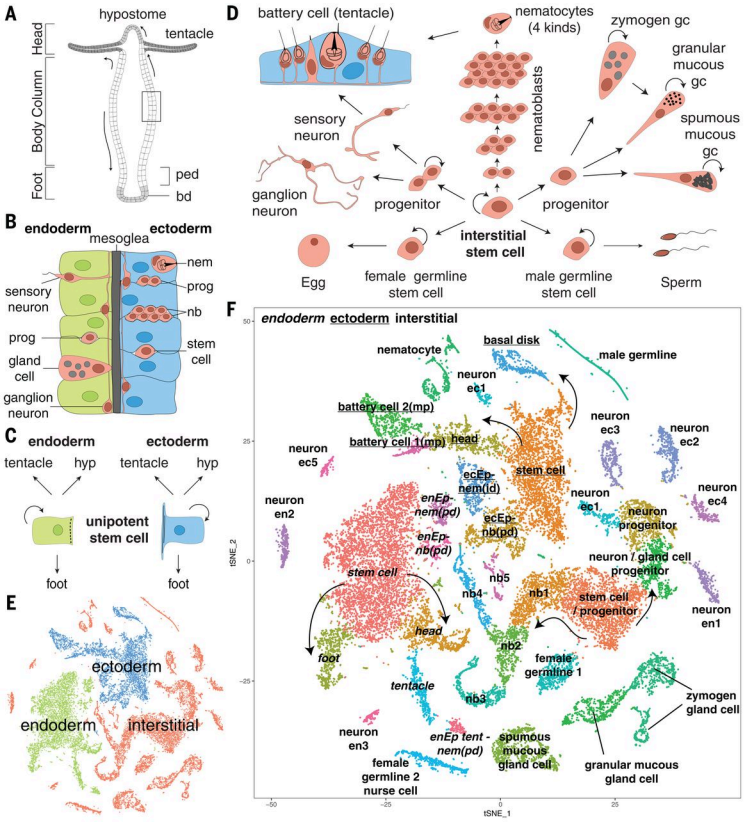


Figure 1.1. *Hydra* tissue composition and single-cell RNA sequencing of 24,985 *Hydra* cells. (A) The *Hydra* body is a hollow tube with an adhesive foot at the aboral end (bd, basal disk; ped, peduncle) and a head with a mouth and a ring of tentacles at the oral end. The mouth opening is at the tip of a cone-shaped protrusion, the hypostome. (B) Enlargement of box in (A). The body column consists of two epithelial layers (endoderm and ectoderm) separated by an extracellular matrix, the mesoglea. Cells of the interstitial cell lineage (red) reside in the interstitial spaces between epithelial cells, except for gland cells, which are integrated into the endodermal epithelium. Ectodermal cells can enclose nerve cells or nematocytes, forming biological doublets. (C) Epithelial cells of the body column are mitotic, have stem cell properties, and give rise to terminally differentiated cells of the hypostome (hyp), tentacles, and foot. (D) Schematic of the interstitial stem cell lineage. The lineage is supported by a multipotent interstitial stem cell (ISC) that gives rise to neurons, gland cells, and nematocytes; ISCs are also capable of replenishing germline stem cells if they are lost. (E) t-SNE representation of clustered cells colored by cell lineage. (F) t-SNE representation of clustered cells annotated with cell state. ec, ectodermal; en, endodermal; Ep, epithelial cell; gc, gland cell; id, integration doublet; mp, multiplet; nb, nematoblast; nem, differentiated nematocyte; pd, suspected phagocytosis doublet; prog, progenitor. id, mp, and pd are categories of biological doublets. Arrows indicate suggested transitions from stem cell populations to differentiated cells. [(A) to (D) adapted from (Siebert, 2018)]

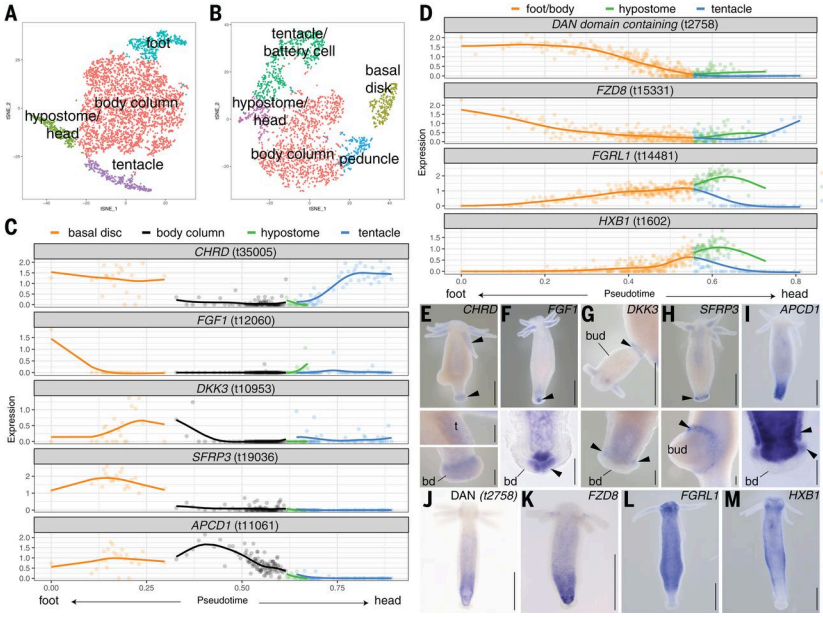


Figure 1.2. Identification of genes with differential expression along the oral-aboral axis. (A and B) t-SNE representation of subclustered endodermal epithelial cells (A) and subclustered ectodermal epithelial cells (B). (Cand D) Epithelial cells were ordered using URD to reconstruct a trajectory where pseudotime represents spatial position. Scaled and log-transformed expression is visualized. (C) Trajectory plots for previously uncharacterized putative signaling genes expressed in ectodermal epithelial cells of foot and tentacles. Genes: BMP antagonist *CHRD* (t35005), *FGF1* (t12060), and Wnt antagonists *DKK3* (t10953), *SFRP3* (t19036), and *APCD1* (t11061). (D) Trajectory plots for genes expressed in a graded manner in endodermal epithelial cells. Genes: BMP antagonist "DAN domain–containing gene" *t2758*, secreted Wnt antagonist *FZD8* (t15331), FGF receptor *FGRL1* (t14481), homeobox protein *HXB1* (t1602). (E to M) Epithelial expression patterns obtained using RNA in situ hybridization consistent with predicted patterns. Whole mounts and selected close-ups are shown. Arrowheads indicate ectodermal signal. t, tentacle; bd, basal disk. Scale bars: whole mounts [including (G)], 500 μm; close-ups, 100 μm.

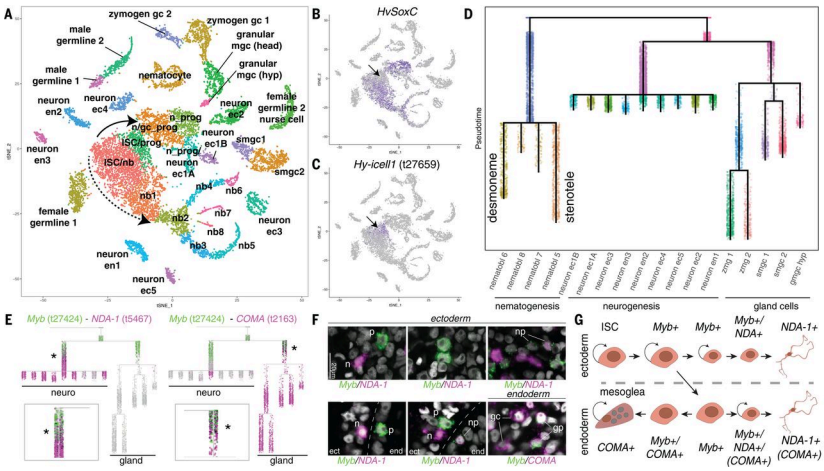


Figure 1.3 Trajectory reconstruction for cells of the interstitial lineage suggests a cell state common to neurogenesis and gland cell differentiation. (A) t-SNE representation of interstitial cells with clusters labeled by cell state. Solid arrow, neurogenesis/gland cell differentiation; dashed arrow, nematogenesis. (B) *HvSoxC* expression in progenitor cells. Arrow indicates putative ISC population, which is negative for *HvSoxC*. (C) The same putative ISC population as in (B) is positive for biomarker *Hy-icell1* expression. (D) URD differentiation tree of the interstitial lineage. Colors represent URD segments and do not correspond to the colors in the t-SNE (see Figure 1.25). (E) *Myb* (green) is expressed in the neuron/gland cell progenitor state and during early neurogenesis/gland cell differentiation. Expression of *Myb* (green, >0) partially overlaps with high expression of the neuronal gene *NDA-1* (magenta, >3) and the gland cell gene *COMA* (t2163) (magenta, >0); *COMA* is also expressed in a subset of endodermal neurons. Coexpressing cells are black. Star and close-up highlights cell states with coexpression. (F) Double labeling using fluorescent RNA in situ hybridization is consistent with neuron differentiation in the endodermal and ectodermal epithelial layers and demonstrates the existence of transition states observed in the trajectory analysis. Additionally, endodermal gland cell differentiation transition states were observed in the endodermal epithelial layer (see also Figure 1.28). gc, gland cell; gp, gland cell progenitor; n, neuron; np, neuron progenitor; p, *Myb*-positive progenitor. (G) Model for progenitor specification. Ectodermal ISCs give rise to a progenitor that can give rise to ectodermal neurons. Progenitors that translocate to the endoderm are able to give rise to gland cells or neurons. ec, ectoderm; en, endoderm; gmgc, granular mucous gland cell; gc, gland cell; hyp, hypostome; ISC, interstitial multipotent stem cell; mgc, mucous gland cell; nb, nematoblast; smgc, spumous mucous gland cell; prog, progenitor; zmg, zymogen gland cell.

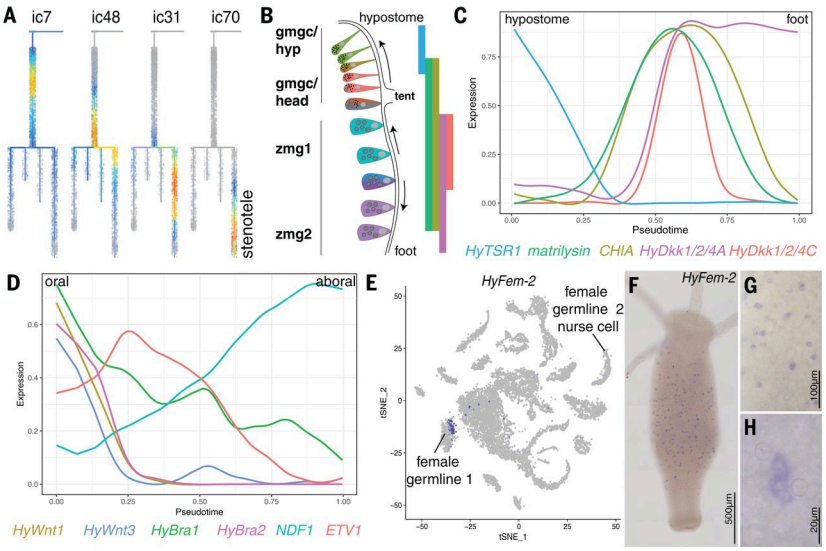


Figure 1.4. Subtrajectory analyses of interstitial cell types. (A) Interstitial gene modules successively expressed in nematocytes forming a stenotele (ic, interstitial gene module). (B) Model for gland cell (ZMG/gMGC) location-dependent changes. Gland cells integrated into the endodermal epithelium get displaced toward the extremities and undergo changes in expression and morphology. Bars show known expression domains for genes depicted in (C). gmgc, granular mucous gland cell; hyp, hypostome; tent, tentacle; zmg, zymogen gland cell. (C) URD linear ZMG/gMGC trajectory recapitulates known position-dependent gene expression in gland cells along the body column. Genes: *HyTSR1* (Siebert et al., 2008), *HyDkk1/2/4A* and *HyDkk1/2/4C* (Augustin et al., 2006; Guder et al., 2006), *matrilysin-like* (t32151), and *CHIA* (t18356) (Figure 1.32 B to E). (D) URD linear sMGC trajectory plot for *HyWnt1* (Lengfeld et al., 2009), *HyWnt3* (Hobmayer et al., 2000), *HyBra1* (Technau and Bode, 1999), *HyBra2* (Bielen et al., 2008), *ETV1* (t22116), and *NDF1* (t21810) showing expression changes in pseudotime that correlate to position along the oral-aboral axis. Cells are ordered according to pseudotime, with putative hypostomal cell states to the left and putative lower head cell states to the right. (E) Plot showing *HyFem-2* expression in a subset of cells in the early female cluster. (F to H) *HyFem-2* is expressed in single cells or pairs scattered within the body column.

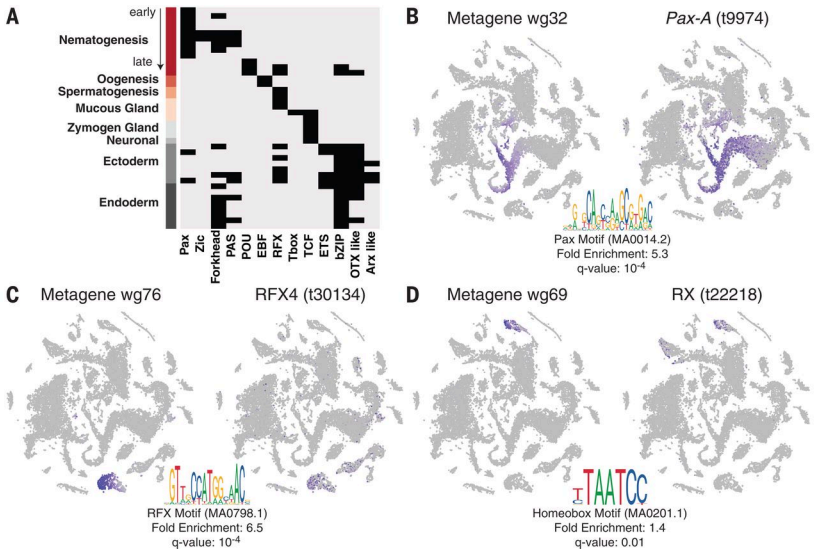


Figure 1.5. Motif enrichment analysis for gene modules and identification of candidate regulators. (A) Selected enriched motifs (columns) found in open chromatin of putative 5′ cis-regulatory regions of coexpressed gene sets (metagenes) for listed cell states (rows). (B to D) Metagene scores visualized on the t-SNE representation (left in each panel), a significantly enriched motif found in putative 5′ cis-regulatory regions (bottom), and candidate regulators likely to bind the identified motif with correlated expression (right). (B) Metagene expressed during nematogenesis and putative PAX regulator. (C) Metagene expressed in gland cells and putative RFX regulator. (D) Metagene expressed in ectodermal epithelial cells of the foot and putative homeobox regulator.

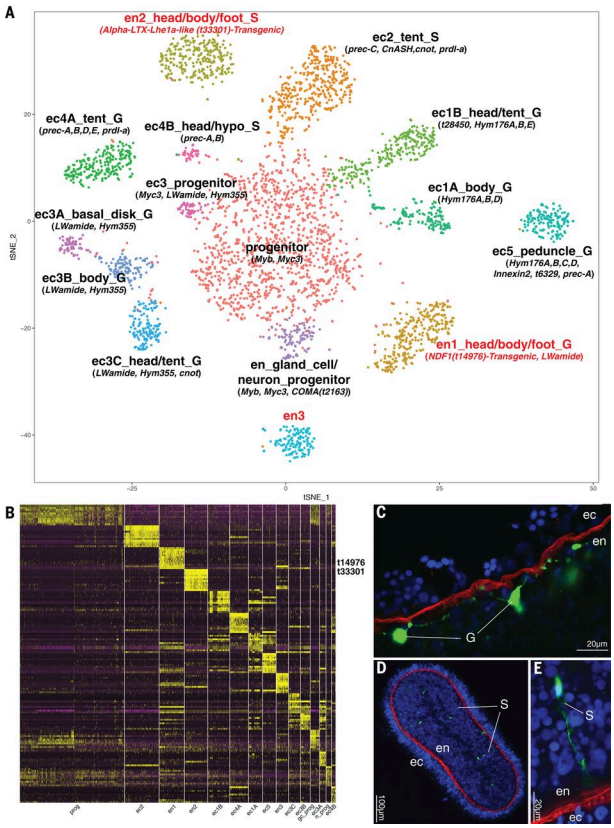


Figure 1.6. Molecular map of the *Hydra* nervous system with spatial resolution.
(A) Subclustering of neurons and neuronal progenitors. Cell states are annotated with cell layer, localization along the body column, tentative neuronal subtype category [sensory (S) or ganglion (G)], and gene markers used in annotations. (B) Heat map shows top 12 markers for neuronal cell states. (C to E) First molecular markers for endodermal neurons. (C) Transgenic line NDF1(t14976)::GFP expressing GFP in endodermal ganglion neurons along the body column (cluster en1). [(D) and (E)] Body column cross section of transgenic line Alpha-LTX-Lhe1a-like(t33301)::GFP expressing GFP in putative sensory neurons (cluster en2). Phalloidin staining (red) marks actin filaments running along the ECM; Hoechst (blue) marks nuclei. en, endoderm; ec, ectoderm.

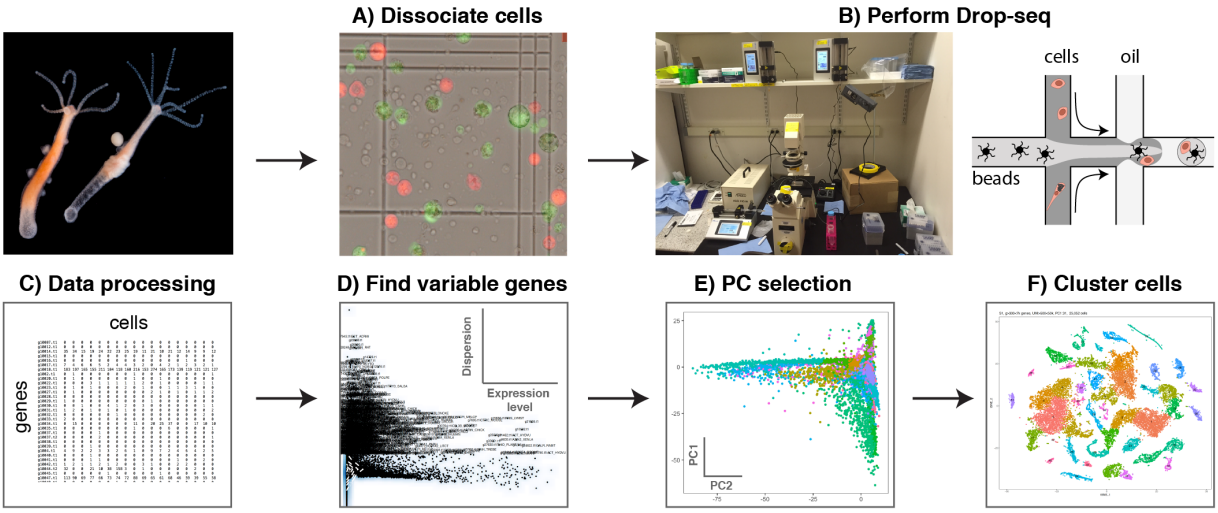


Figure 1.7. Drop-seq workflow. (A) *Hydra vulgaris* AEP polyps were dissociated into single cells using Pronase E for 1.5h (*57*). (B) Single cell transcriptomes were resolved by droplet-based sequencing (Drop-seq), which encapsulates single cells in nanoliter oil droplets, lyses cells within those droplets, and captures polyadenylated transcripts with barcoded oligo(dT) covalently linked to beads (Macosko et al., 2015). The barcodes facilitate cell assignment ('STAMP' barcodes) and elimination of library amplification artifacts using unique molecular identifiers (UMIs). Droplet generation was followed by reverse transcription, PCR amplification, and Nextera library preparation. Libraries were sequenced on an Illumina NextSeq 500 sequencer with a custom sequencing strategy. Raw reads were quality filtered and aligned to a *Hydra vulgaris* AEP transcriptome and the *Hydra vulgaris* (strain 105) 2.0 genome reference. (C) A digital expression matrix was constructed using Drop-seq tools v1-2.12 (*http://mccarrolllab.com/dropseq/*). 15 Drop-seq runs resulted in a total of 24,985 sequenced *Hydra* cells after initial filtering. (D-E) We used Seurat (Satija et al., 2015) for the identification of genes with relatively high average expression and variability, and subsequent principal component analysis. (F) Principal components were used in graph-based clustering. Clusters were visualized on t-distributed stochastic neighbor embedding (t-SNE) plots.

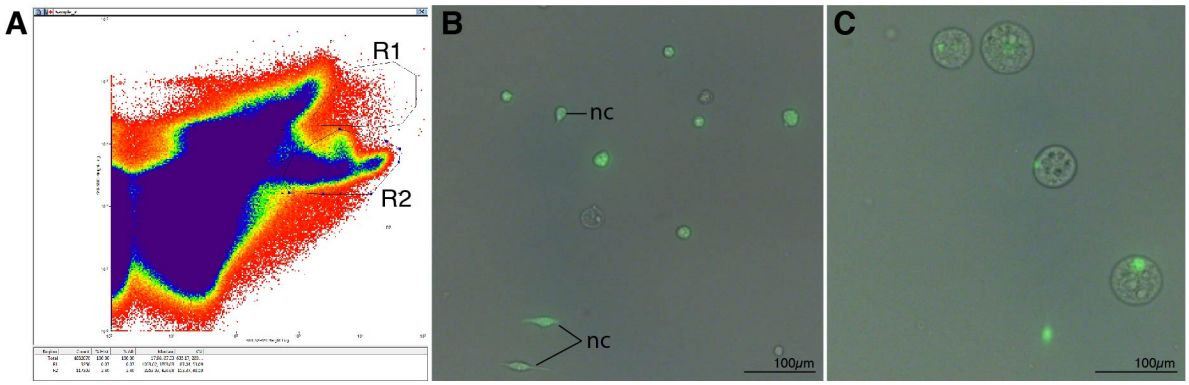


Figure 1.8. Fluorescence-activated cell sorting (FACS) of *Hydra* cell suspensions. (A) Gates implemented to collect GFP-positive cells. We used transgenic *Hydra vulgaris* AEP line nGreen, which expresses GFP driven by an actin promoter predominantly in neuronal cells. Both populations, R1 and R2, are GFP- positive. Population R2 (B) contained undifferentiated cells, neurons (nc), and additional cells of the neuronal lineage (see Figure 1.49). Cells of population R2 were collected in two independent sortings and used to generate two Drop-seq libraries (libraries 12-N1, 12-N2). (C) Population R1 was characterized by lower GFP intensities and larger cell sizes. Fluorescent microscopy combined with DIC (C) revealed GFP signal within larger epithelial cells demonstrating the existence of multiplets likely due either to phagocytosis or to naturally existing cell doublets; *Hydra* epithelial cells can house other cell types such as nematocytes and neurons (see Materials and Methods).

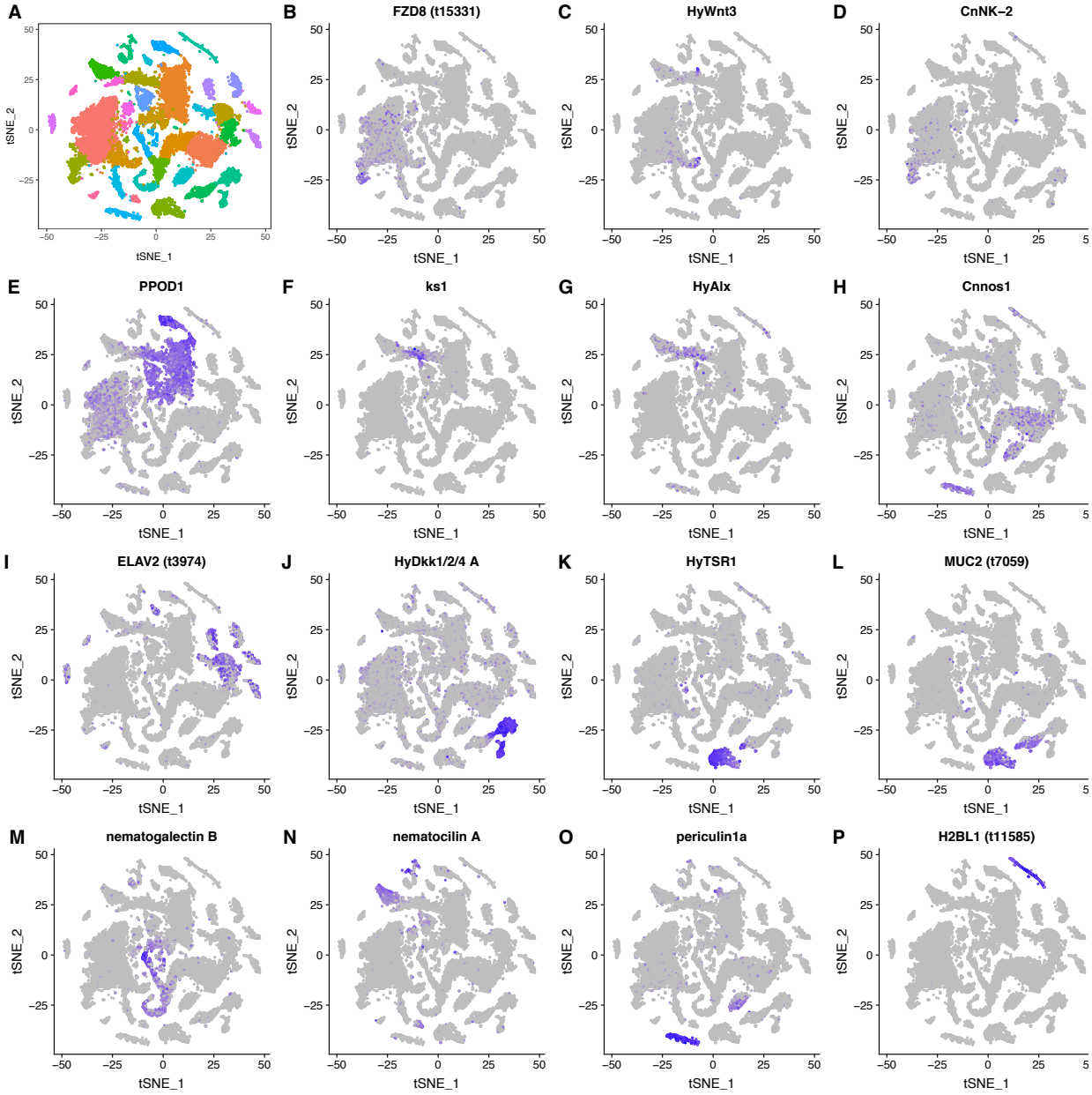


Figure 1.9. Cluster annotation. Selected markers used for cluster annotation. (A) t-SNE representation of the clustering. For annotated clusters see Figure 1.1 F. (B) Endoderm - *FZD8* (t15331) (this study, Figures 1.2 K, 1.18 C, and 1.19). (C) Endoderm/ectoderm hypostome - *HyWnt3* (Hobmayer et al., 2000). (D) Endoderm foot/peduncle - *CnNK-2* (Grens et al., 1996). (E) Ectoderm/endoderm - *PPOD1* (Thomsen and Bosch, 2006). (F) Ectoderm head - *ks1* (Endl et al., 1999). (G) Ectoderm tentacle - *HyAlx* (Smith et al., 2000). H) Multipotent i-cells/progenitors/female germline - *Cnnos1*. (I) Neuron progenitor/neurons - *ELAV2* (t3974) (this study, Figure 1.46 L-M). (J) Zymogen gland cell - *HyDkk1/2/4 A* (Augustin et al., 2006; Guder et al., 2006). K) Mucous gland cells - *HyTSR1* (Siebert et al., 2008). (L) Granular and spumous mucous gland cells - *MUC2* (t7059). *(*M) Nematogenesis/biological doublets - *nematogalectin B* (Hwang et al., 2010). N) differentiated nematocytes/battery cell - *nematocilin A* (Hwang et al., 2008). (O) female germline - *periculin1a* (Fraune et al., 2010). (P) male germline - *H2BL1* (t11585) (this study, Figure 1.36 A).

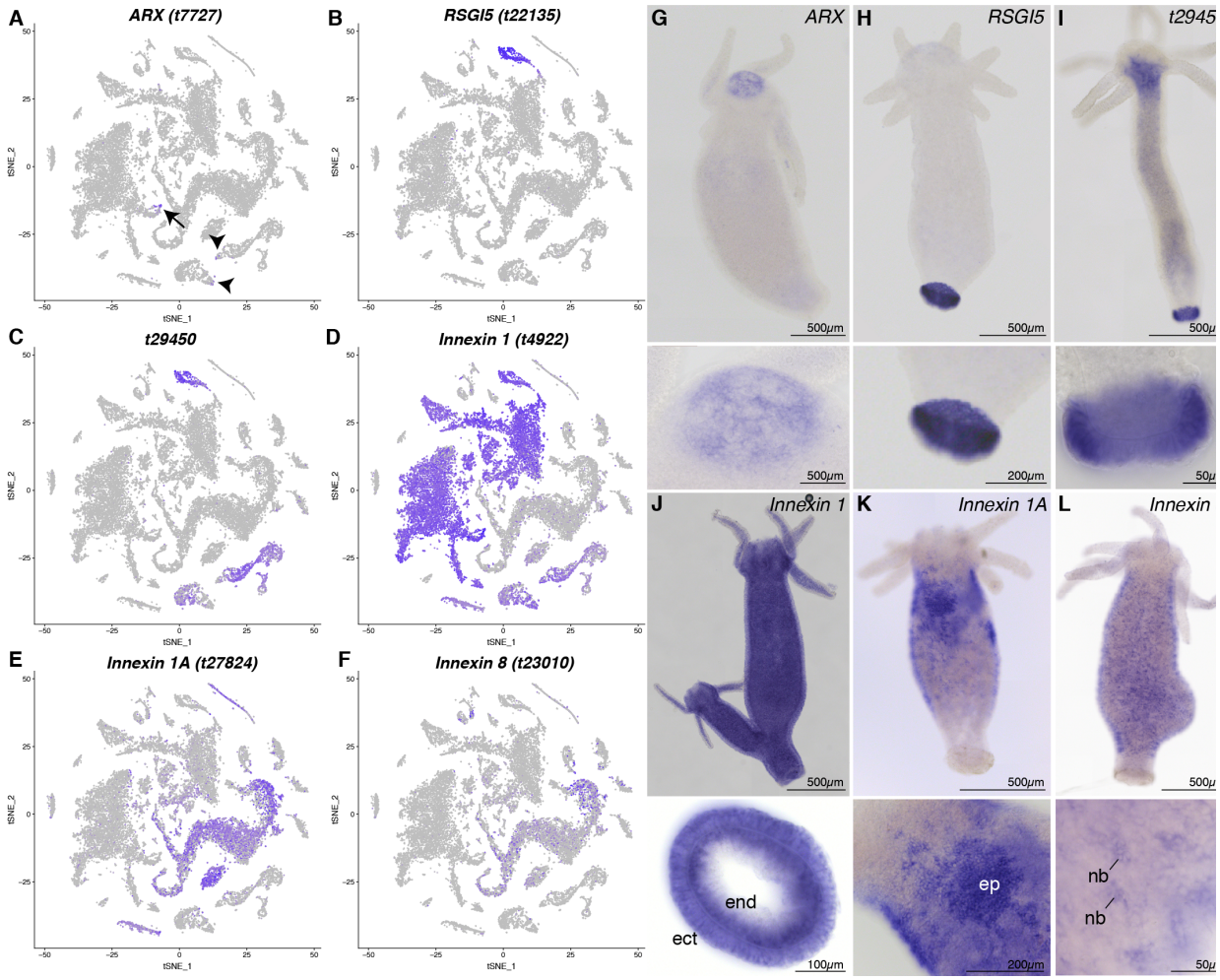


Figure 1.10. Examples of cluster biomarker validation. (A-F) t-SNE plots for six genes expressed in specific cell clusters. (G-N). Predicted expression patterns are validated by in situ hybridizations. Presented as whole mounts and close-ups. (A) Homeobox gene *ARX* (t7727) is expressed in epithelial (arrow) and gland cells (arrowheads) of the hypostome. (B) *RSGI5* (22135) is expressed in ectodermal basal disk cells. (C) t29450 is expressed in ectodermal basal disk cells and both granular and spumous gland cells. (D) *Innexin 1* (t4922) is expressed in endodermal and ectodermal epithelial cells (Alexopoulos et al., 2004). (E) *Innexin 1A* (t27824) expression in male and female germ line cells, interstitial stem cells and progenitors. (F) *Innexin 8* (t23010) expression in differentiating progenitor cells. (G) *ARX* (t7727) expression in hypostomal cells. (H) *RSGI5* (t22135) expression in ectodermal basal disk cells. (I) *t29450* expression in endodermal gland cells and ectodermal basal disk cells. (J) *Innexin 1* expression throughout the body column. Close-up of a body column cross-section reveals expression in ectodermal (ect) and endodermal (end) epithelial cells. (K) *Innexin 1A* expression in interstitial cells along the body column. Close-up shows expression in a developing egg patch (ep). (L) *Innexin 8* with expression in interstitial cells throughout the body column. Close-up shows nematoblast nests (nb).

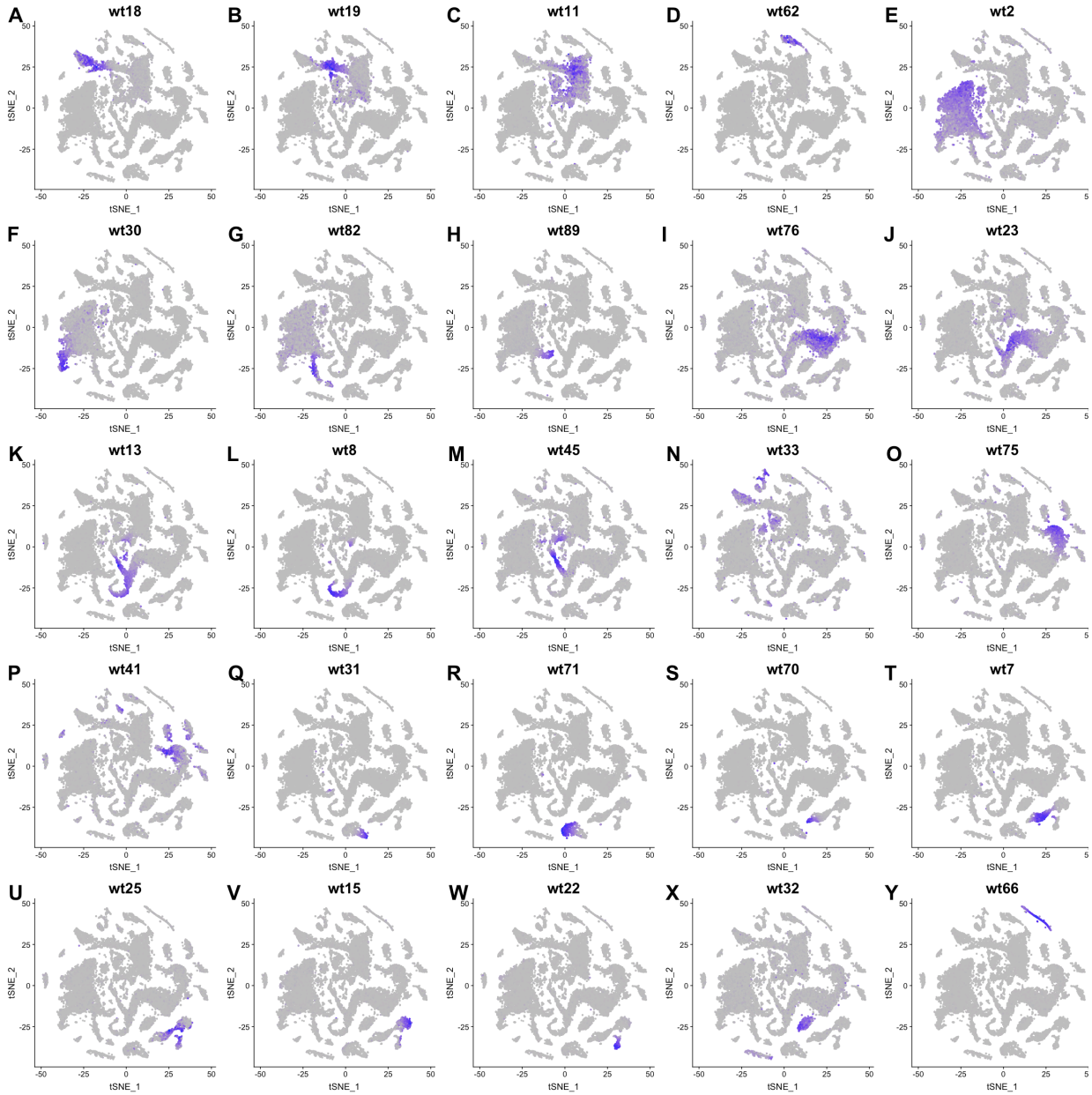


Figure 1.11. Selected metagenes identified in NMF analysis for the whole dataset (NMF analysis wt_K96). Metagenes are groups of genes with similar expression patterns, as identified by non-negative matrix factorization (NMF). Here, the overall expression of each metagene is displayed. For annotated clusters see Figure 1.1 F. (A) Tentacle ectodermal epithelial cells. This metagene includes transcripts that are expressed in the epithelial cell of a battery cell complex since expression is not found in neuronal or nematocyte cell populations. (B) Ectodermal epithelial cells, head. (C) Ectodermal epithelial cells, body column. (D) Ectodermal epithelial cells, basal disk. (E) Endodermal epithelial cells, body column. (F) Endodermal epithelial cells, foot. (G) Endodermal epithelial cells, tentacle. (H) Endodermal epithelial cells, hypostome. (I) Interstitial stem cells and early progenitors. (J) Early stage nematoblast, singletons and phagocytosed. (K) Mid stage nematoblast, singletons and phagocytosed. L,M) Late nematoblast, singletons and integrated into a battery cell. (N) Mature nematocyte, singletons and integrated. (O) Neuronal cell progenitors. (P) Differentiated neurons. (Q). Spumous mucous gland cells,

hypostome. (R) Spumous mucous gland cells, mid/lower head. (S) Granular mucous gland cells, hypostome. (T) Granular mucous gland cells, mid/lower head. (U) Granular mucous gland cells/zymogen gland cells. (V,W) Zymogen gland cells. (X) Female germline cells. (Y) Male germline cells.

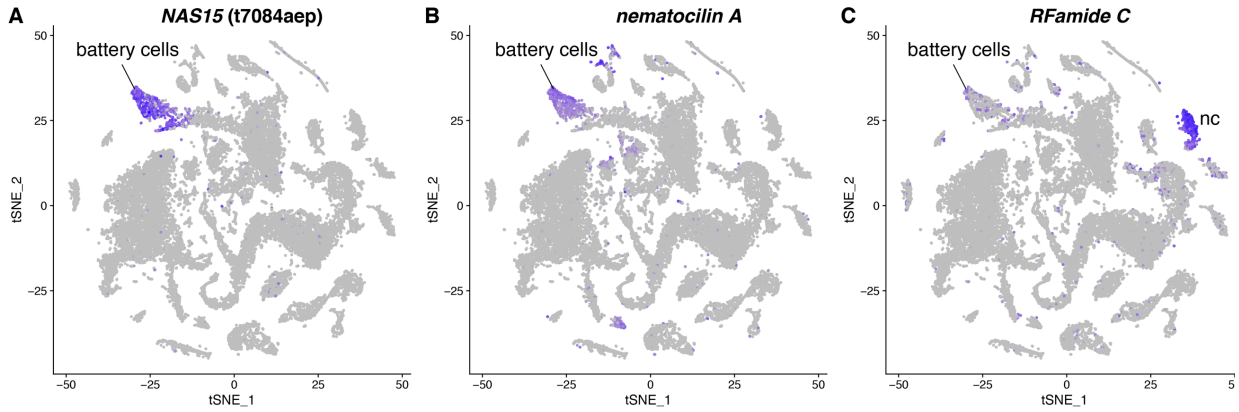


Figure 1.12. Co-expression of epithelial, nematocyte, and neuronal markers in multiplets of the battery cell cluster. Battery cells are epithelial cells in the tentacle ectoderm that house multiple nematocytes and a neuron. These are an example of a biological doublet, where cells that are tightly associated in the animal and cannot be dissociated are repeatedly sequenced as hybrid transcriptomes. (A) Metalloproteinase *NAS15* (t7084) is a high scoring gene within metagene wt18 (NMF analysis wt_K96). This metagene represents a set of genes expressed in the epithelial cell of a battery cell complex. (B) *Nematocilin A* (t23176) is expressed in mature nematocytes. The *nematocilin A* protein is localized in the central filament of the mechanosensory cilium (Hwang et al., 2008). (C) *RFamide preprohormone C* (t25706) is expressed in neurons of the tentacles (Darmer et al., 1998). nc: neuron cluster.

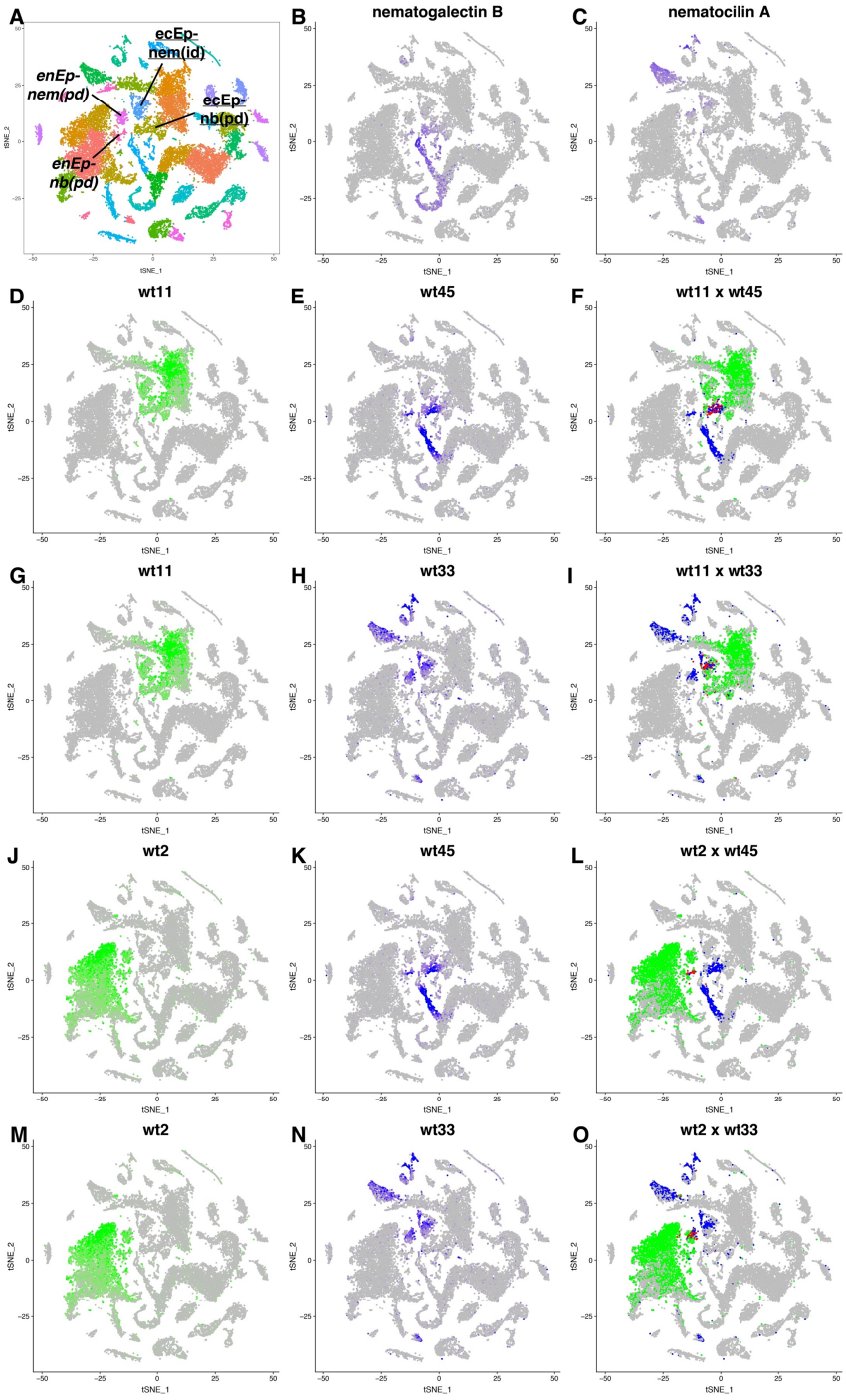


Figure 1.13. Co-expression analyses identify biological multiplets and suspected phagocytosis doublets. Gene modules describe a particular process or cell state in the data set. We used known gene expression patterns to annotate NMF gene modules and classified cells that express unexpected combinations of gene modules as doublets. This strategy was implemented in URD and used for doublet removal prior to trajectory reconstruction. (A) t-SNE with annotated doublet clusters involving epithelial cells of the body column. Cells from these doublet clusters were not considered in subclusterings of cells from different cell lineages. For complete cluster annotations see Figure 1.1 F. (B) Nematoblast gene *nematogalectin B* (Hwang

et al., 2010) highlights metagene wt45 as a nematoblast metagene. (C) *Nematocilin A* (Hwang et al., 2008) is expressed in mature nematocytes and highlights wt33 as a nematocyte metagene. D,G) Metagene wt11 is expressed in ectodermal epithelial cells of the body column. (D-F) Co-expression (red cells) of metagene wt11 (green cells) and wt45 (blue cells) suggests phagocytosed nematoblasts in ectodermal epithelial cells of cluster ecEp-nb(pd). (G-I) Co-expression (red cells) of metagene wt11 (green cells) and wt33 (blue cells) suggests integrated/mounted nematocytes in ectodermal epithelial cells of cluster ecEp-nem(id). (J,M) Metagene wt2 (green cells) is expressed in endodermal epithelial cells of the body column. (J-L) Co-expression (red cells) of metagenes wt2 (green cells) and wt45 (blue cells) suggests phagocytosed nematoblasts in endodermal epithelial cells of cluster enEp-nb(pd). (M-Q) Co-expression (red cells) of metagenes wt2 (green cells) and wt33 (blue cells) suggests phagocytosed nematocytes in endodermal epithelial cells of cluster enEp- nem(pd). ec: ectodermal, en: endodermal, Ep: epithelial cell, id: integration doublet, nb: nematoblast, nem: mature nematocyte, pd: phagocytosis doublet.

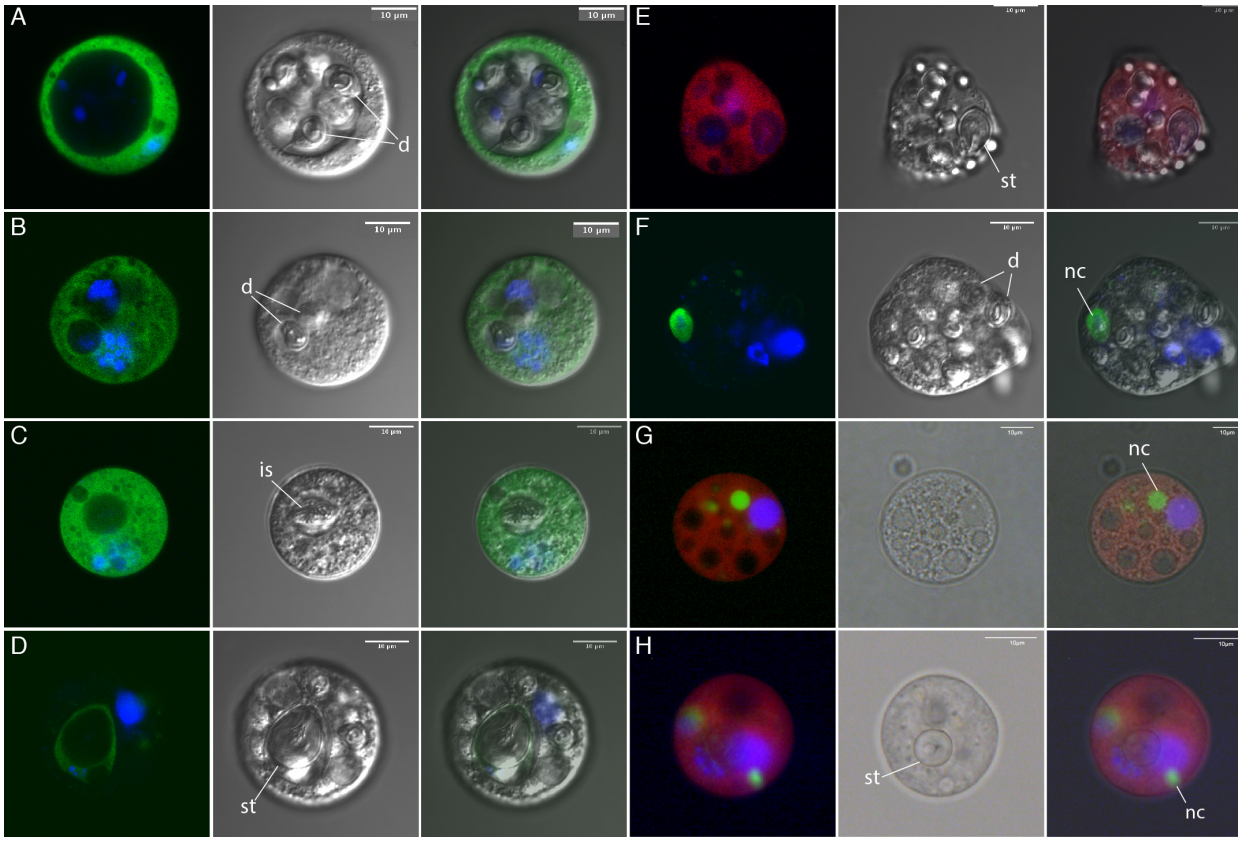


Figure 1.14. Documentation of live cells after tissue dissociation reveals cell multiplets. Multiple transgenic lines expressing fluorescent proteins were used to demonstrate the existence of biological multiplets (e.g. battery cells), viability of cells within the host cells, and the occurrence of phagocytosis. For all panels: 1st image fluorescence, 2nd image bright field, 3rd image overlay. (A-C,E) Line expressing GFP in ectodermal epithelial cells and RFP in endodermal epithelial cells (Glauber et al., 2013). (D,F) Line nGreen expressing GFP predominantly in the neuronal lineage and with scattered expression in nematocytes. (G,H) Line PT1 (courtesy of Rob Steele), cross expressing GFP in ectodermal neurons and DsRed2 in epithelial cells (based on line hym176B::GFP (Noro et al., 2019) and line all DsRed2 (Glauber et al., 2013)). (A-D) Ectodermal epithelial cells containing a single or multiple nematocytes of one or multiple kinds. A) Ectodermal epithelial cell containing multiple nematocytes with desmonemes. The nematocytes appear to be contained in a vacuole like compartment that may be indicative of phagocytotic uptake prior to or in the course of the dissociation procedure. (B) Ectodermal epithelial cell containing two nematocytes with desmonemes. (C) Ectodermal epithelial cell containing a single nematocyte with isorhiza nematocyst. (D) Nematocyte with stenotele and GFP positive cytoplasm within an ectodermal epithelial battery cell. (E) Endodermal epithelial cell containing a stenotele suggesting nematocyte uptake. (F) Ectodermal epithelial battery cell containing a GFP-positive neuron. (G) Endodermal epithelial cell (as indicated by the presence of vacuoles) containing GFP, with transgenic neurons as the sole possible source for GFP. (H) Ectodermal epithelial cell containing a nematocyte (stenotele) and GFP positive structures, with transgenic neurons as the sole possible source for GFP. d: desmoneme, is: isorhiza, st: stenotele, nc: neuronal cell.

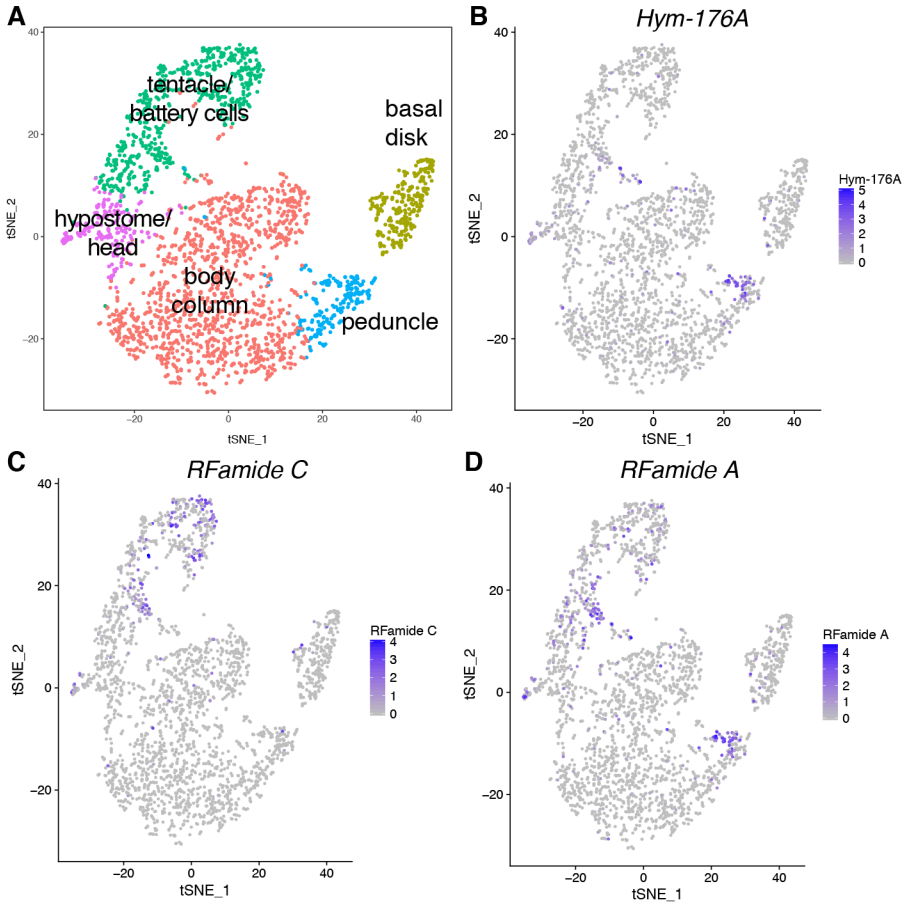


Figure 1.15. Neuronal gene expression in epithelial subclusterings. Subclusterings of epithelial cells (Figure 1.2 A-B) contain cells with doublet signatures partially caused by suspected phagocytic events (Figure 1.14) or cells that are integrated in epithelial cells in homeostatic *Hydra* (e.g. neurons within battery cells) (Figure 1.12). Epithelial clusterings are biologically meaningful despite contaminating expression of interstitial genes, e.g. neuronal genes, suggesting that clusterings are driven by epithelial genes. Doublets were further removed prior to downstream trajectory reconstruction using URD. (A) t-SNE plot for subset of ectodermal epithelial cells. (B) Expression plot for neuropeptide *Hym-176A* reveals reported peduncle, body column and head expression (Yum et al., 1998). (C) Expression plot for neuropeptide *RFamide C* reveals reported tentacle battery cell expression (Darmer et al., 1998). (D) Expression plot for neuropeptide *RFamide A* reveals reported tentacle and basal disk expression (Darmer et al., 1998).

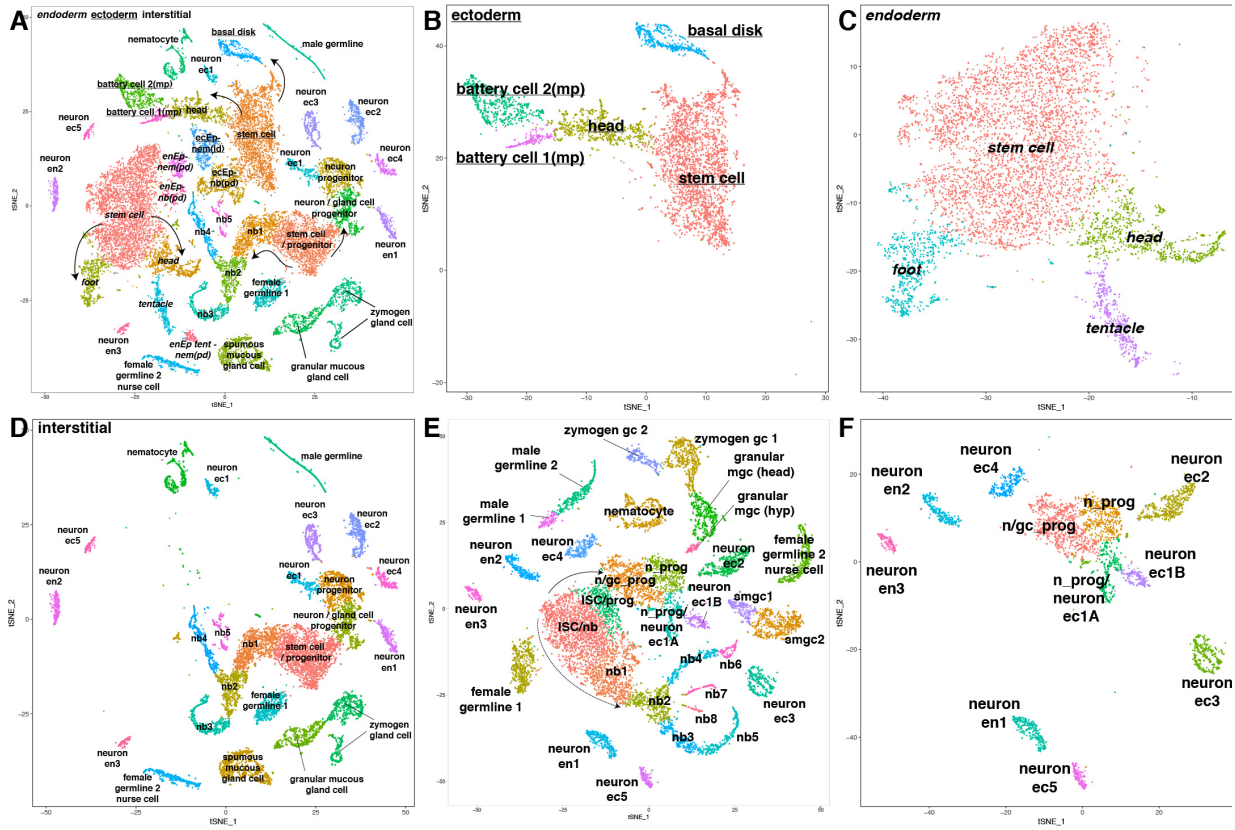


Figure 1.16. Cells included in lineage and neuronal subclusterings. A) t-SNE representation for clustering of all cells in the transcriptome data set (see also Figure 1.1 F). B) Ectodermal epithelial cells extracted for subcluster analysis. We identified library-specific (batch) effects in this cell population and only two sets of libraries were further considered in downstream analyses (see supplementary analysis SA03). The subclustering result is shown in Figure 1.2 B. C) Endodermal epithelial cells extracted for subcluster analysis. The subclustering result is shown in Figure 1.2 A. D) Cells of the interstitial lineage that were extracted for subcluster analysis. The subclustering result is shown in (E). E) Subclustering of interstitial cells (see also Figure 1.3 A). F) Neuronal cells from the interstitial subclustering shown in (E) extracted for neuron subcluster analysis. The subclustering result is shown in Figure 1.6 A. ecEP: ectodermal epithelial cell, enEP: endodermal epithelial cell, gc: gland cell, hyp: hypostome, id: integration doublet, ISC: interstitial stem cell, mgc: mucous gland cell, mp: multiplet, nb: nematoblast, nem: nematocyte, pd: suspected phagocytosis doublet, prog: progenitor, smgc: spumous mucous gland cell, tent: tentacle, zmg: zymogen gland cell.

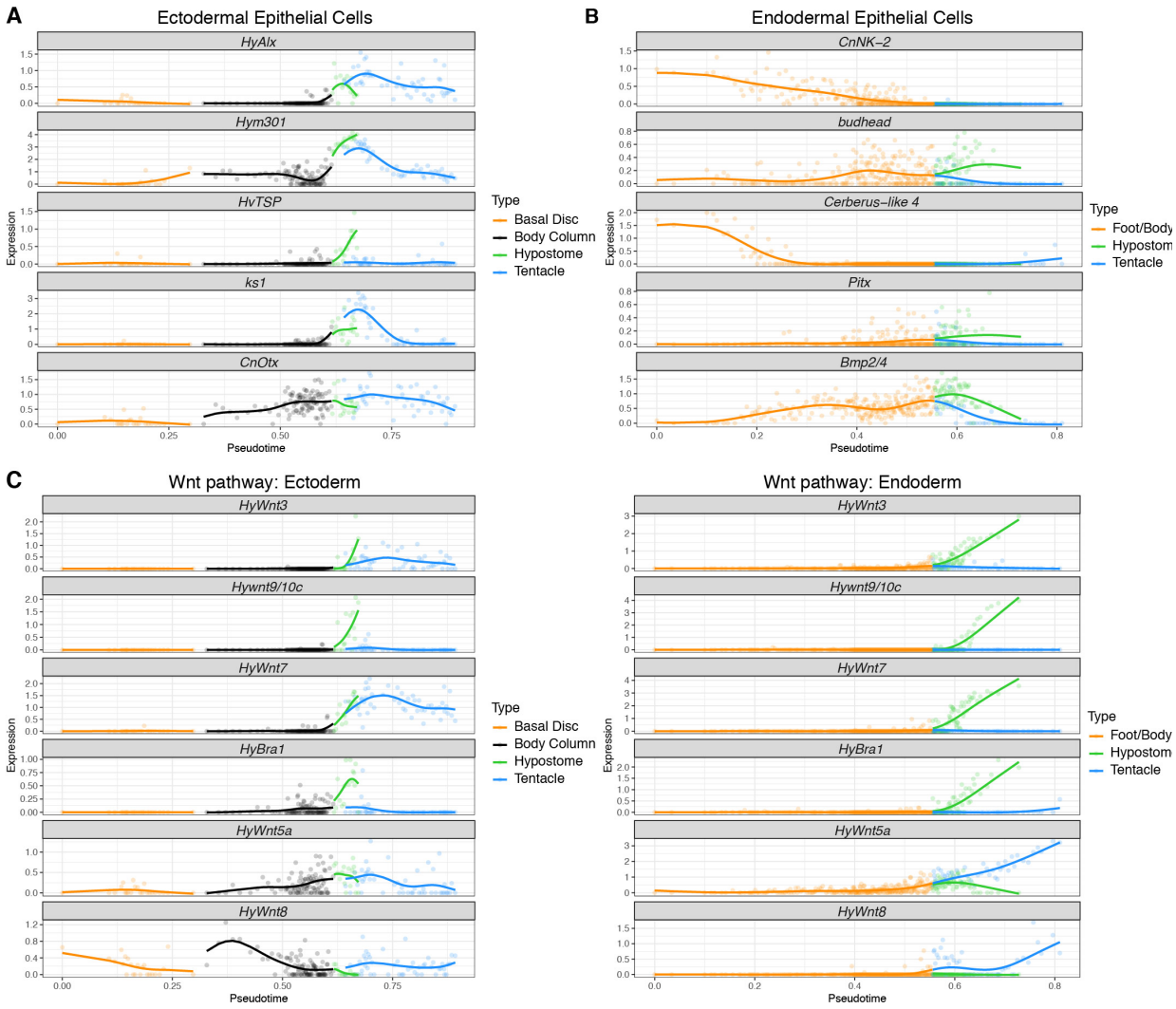


Figure 1.17. Trajectory plots for previously characterized ectodermal and endodermal epithelial genes. Epithelial cells were ordered using URD to reconstruct a trajectory where pseudotime represents spatial position by using foot cells as the root and head cells as the terminal points in the reconstruction. Scaled and log-transformed expression is visualized. Known expression of genes was used to validate trajectories and was recapitulated in the plots. A) Trajectory plots for genes expressed in ectodermal epithelial cells. *HyAlx* (Smith et al., 2000), *Hym301* (Takahashi et al., 2005), *HvTSP* (Lommel et al., 2018), *ks1* (Weinziger et al., 1994) and *CnOtx* (Smith et al., 1999). B) Trajectory plots for genes expressed in endodermal epithelial cells. *CnNK-2* (Grens et al., 1996), *budhead* (Martinez et al., 1997), *Cerberus-like 4* (Watanabe et al., 2014), *Pitx* (Watanabe et al., 2014) and *Bmp2/4* (Watanabe et al., 2014). C) Expression of *Wnt* and *Wnt* downstream genes in ectodermal and endodermal epithelial cells. Canonical Wnt genes *HyWnt3* (Hobmayer et al., 2000), *HyWnt9/10c* (Lengfeld et al., 2009) and *HyWnt7* (Lengfeld et al., 2009), Wnt down-stream target *HyBra1* (Technau and Bode, 1999) and non-canonical Wnt genes *HyWnt5a* and *HyWnt8* (Philipp et al., 2009).

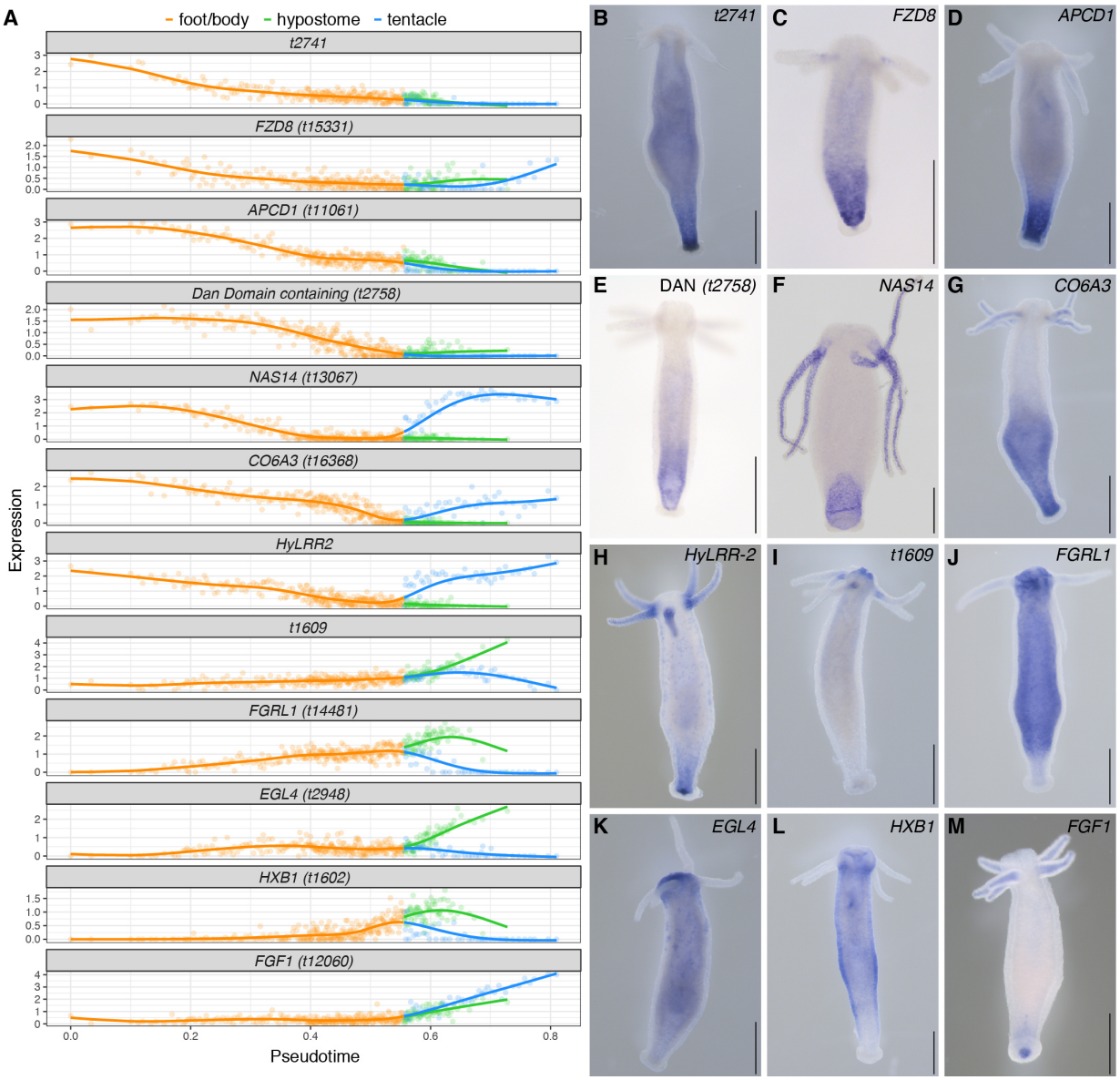


Figure 1.18. Spatially resolved gene expression in endodermal epithelial cells along the body. Epithelial cells were ordered using URD to reconstruct a trajectory where pseudotime represents spatial position by using foot cells as the root and head cells as the terminal points in the reconstruction. Scaled and log- transformed expression is visualized. (A) Expression of a few genes identified as spatially varying from the reconstructed trajectory. Gene selection was focused on graded expression changes from foot to head, head to foot, and within tentacles. (B-D) Epithelial expression patterns were validated using RNA in situ hybridization. The RNA in situ hybridization patterns indicate that cells from the foot and peduncle have pseudotime values from 0 to 0.4 (e.g. compare trajectory and in situ pattern for *NAS14*), while values 0.4 to 0.55 represent the body column. (B) Transcript *t2741*. (C) *FZD8* (t15331). (D) *APCD1* (t11061). (E) Dan domain containing transcript *t2758*, identified using PFAM annotations. (F) *NAS14* (t13067). (G) CO6A3 (t16368). (H) *HyLRR-2* (t18862). This gene has additional expression in cells of the interstitial lineage (subset of neurons). (I) Transcript *t1609*. (J) *FGRL1* (t14481). (K)

EGL4 (t2948). This gene has additional expression in cells of the interstitial lineage (neurons, subset of gland cells). (L) *HXB1* (t1602). M) *FGF1* (t12060).

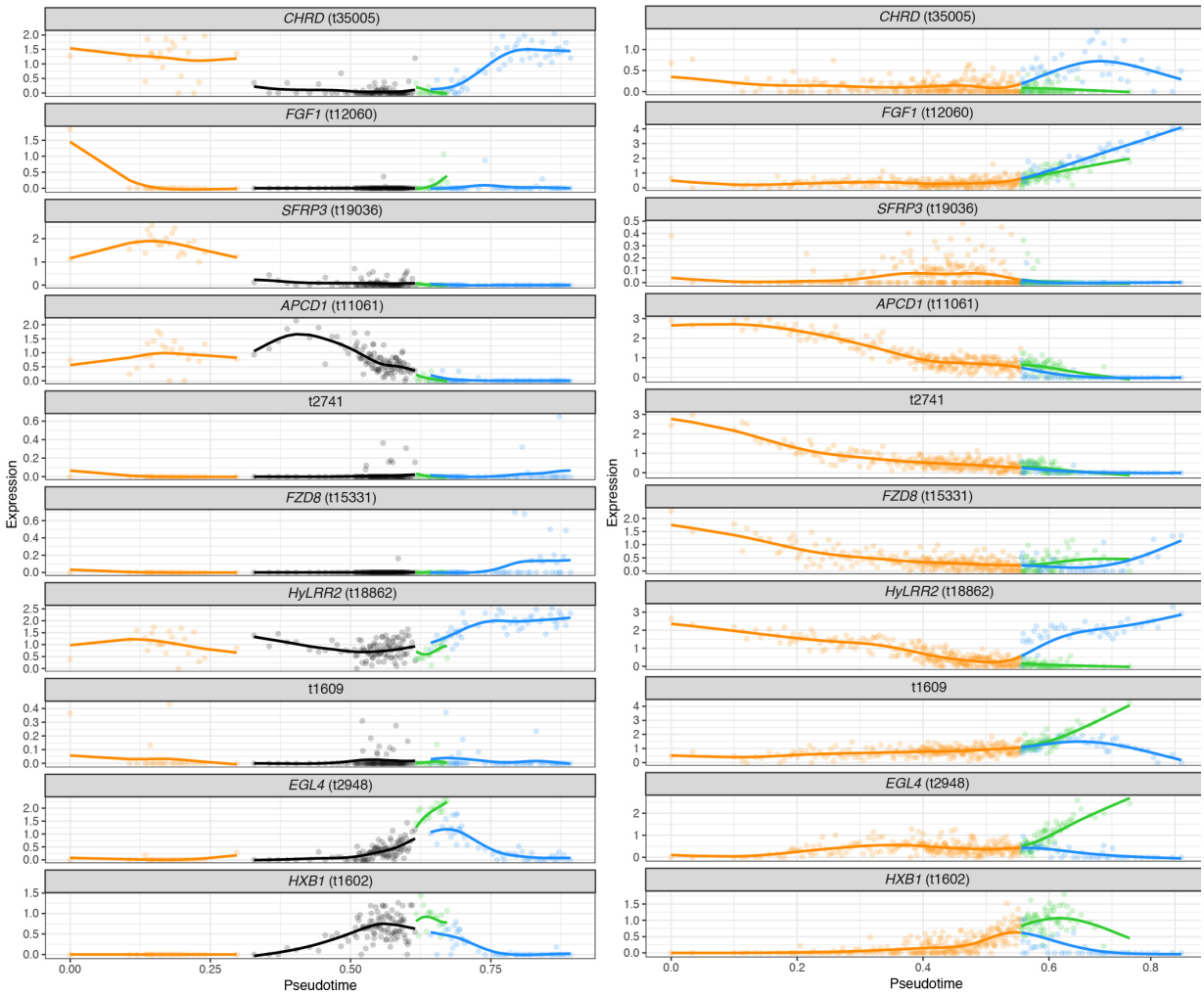
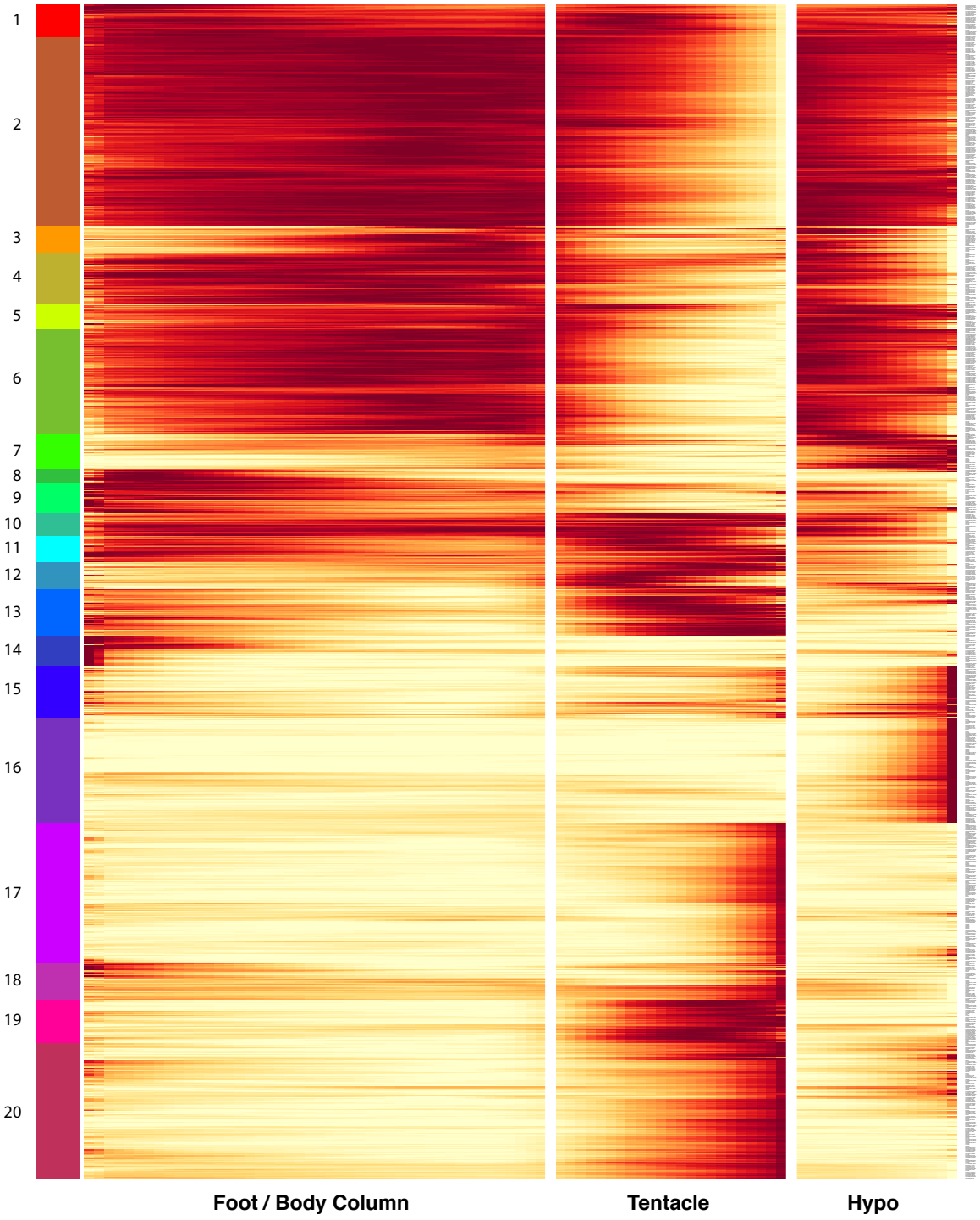
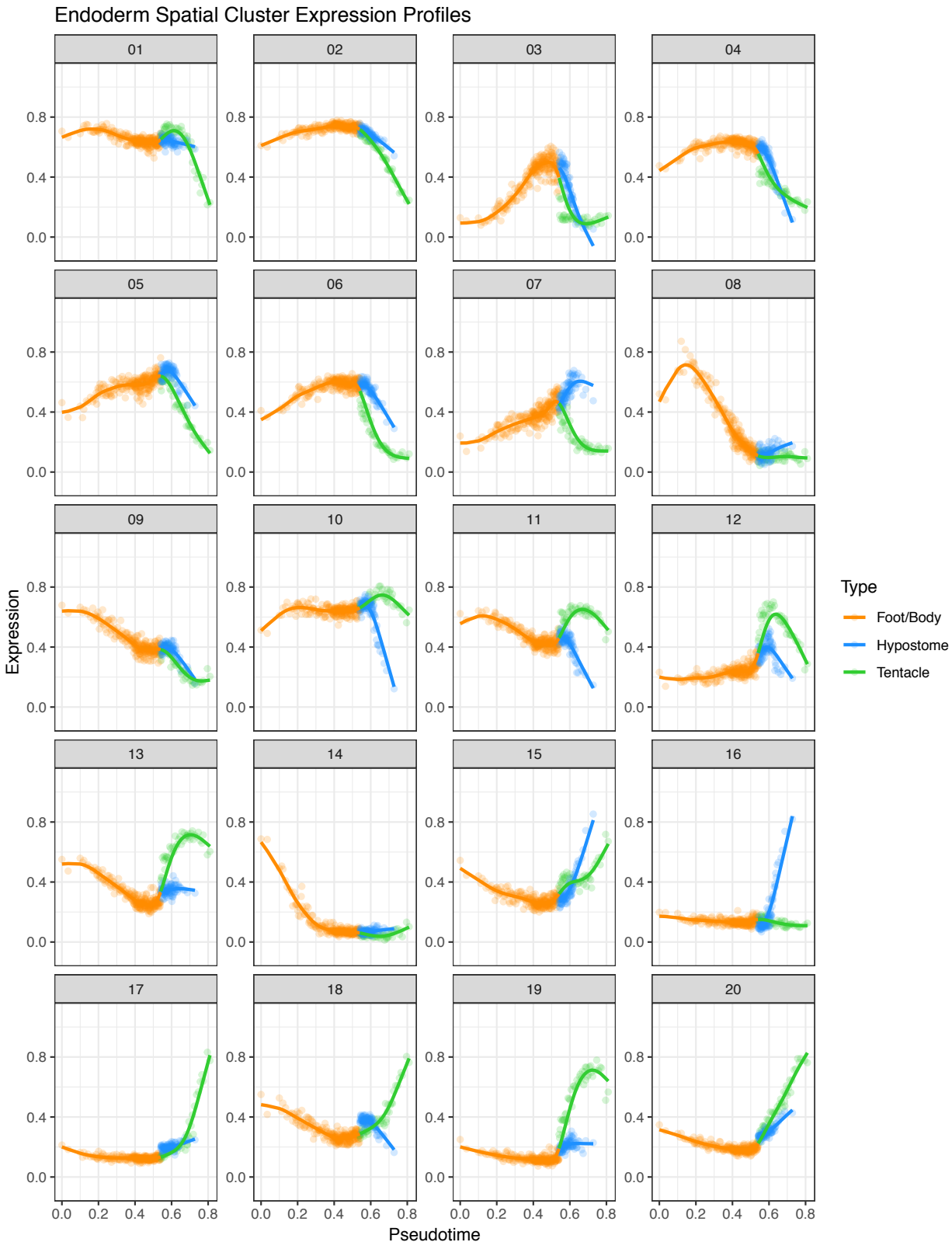


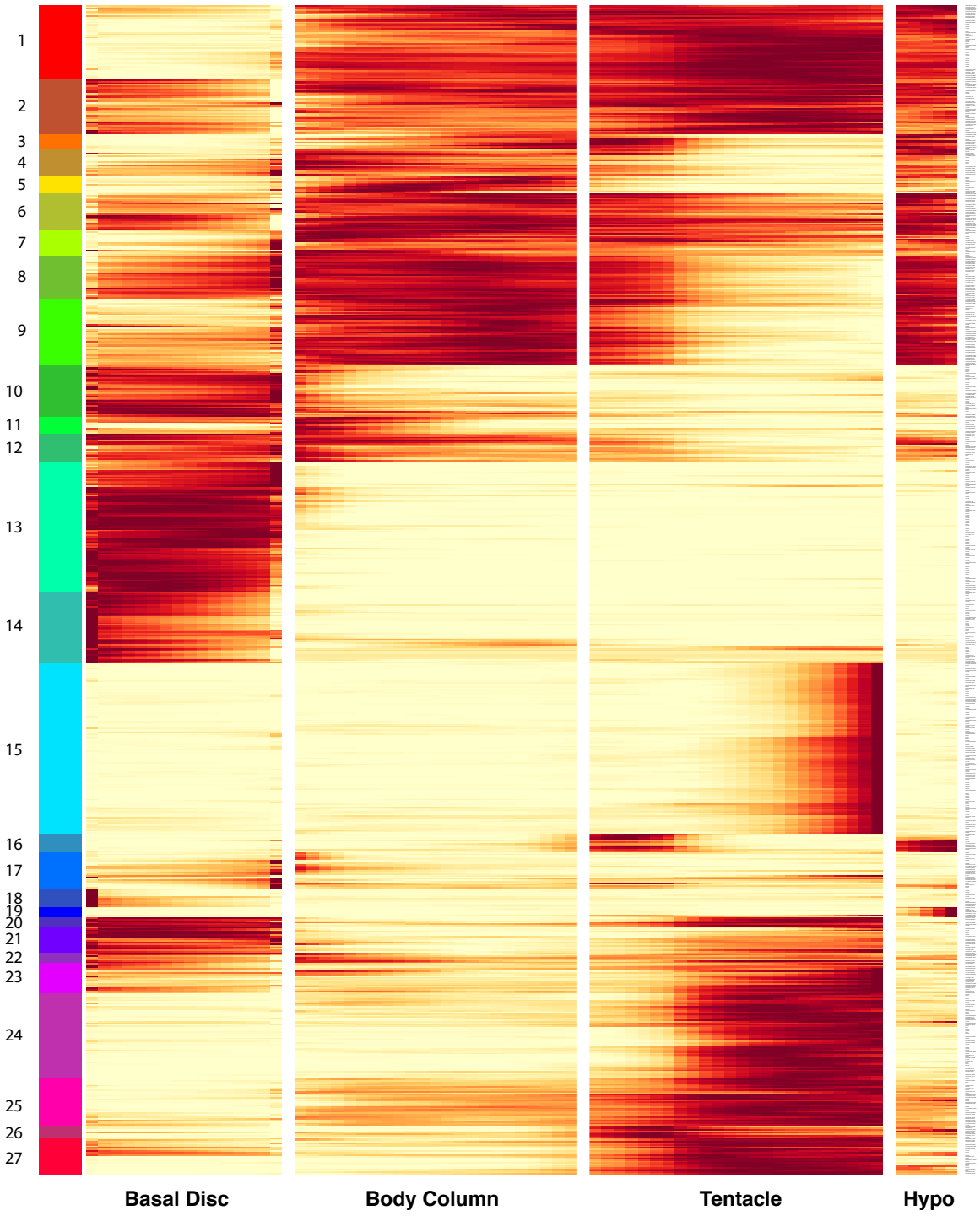
Figure 1.19. Trajectory plots for selected genes with detected expression in both endodermal and ectodermal epithelial cells. Set of genes featured in Figure 1.2 and 1.18 with endodermal and ectodermal expression presented side by side. We find that the stronger signal in one tissue can obfuscate a signal from the other tissue when applying colorimetric visualization approaches.

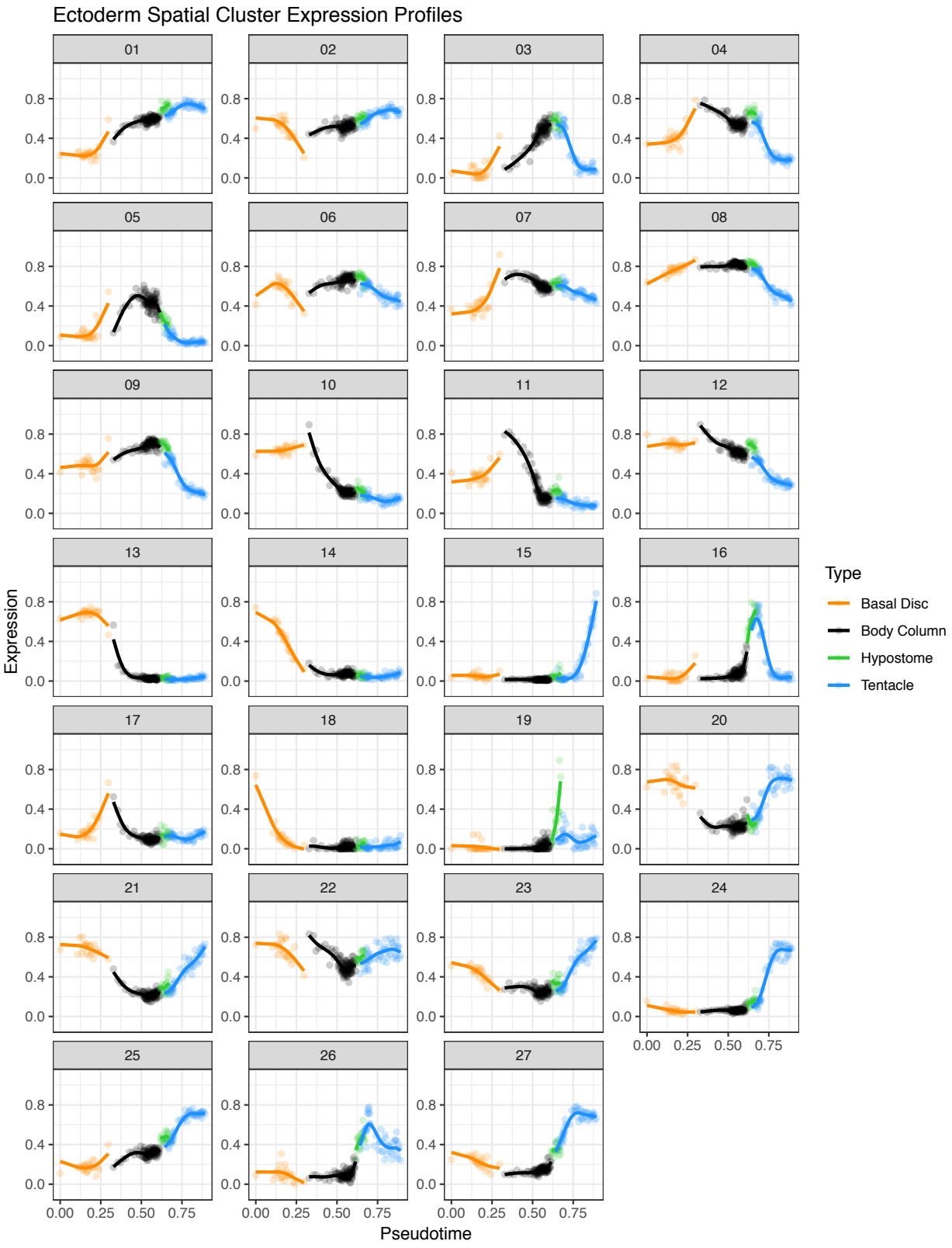
Endoderm Spatial Expression



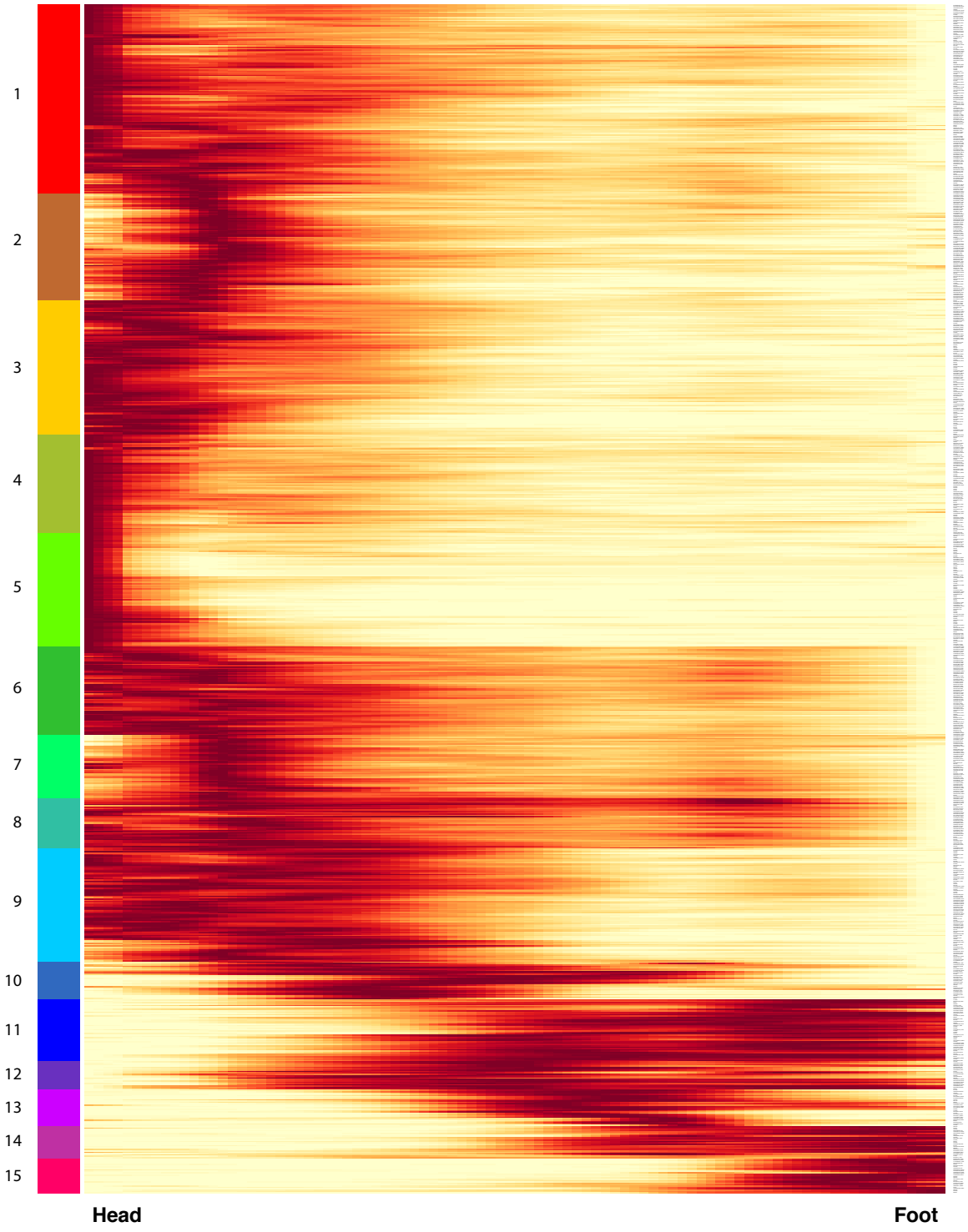

Endoderm Spatial Cluster Expression Profiles
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
Expression
Pseudotime
0.0
0.2
0.4
0.6
0.8
Type
Foot/Body
Hypostome
Tentacle

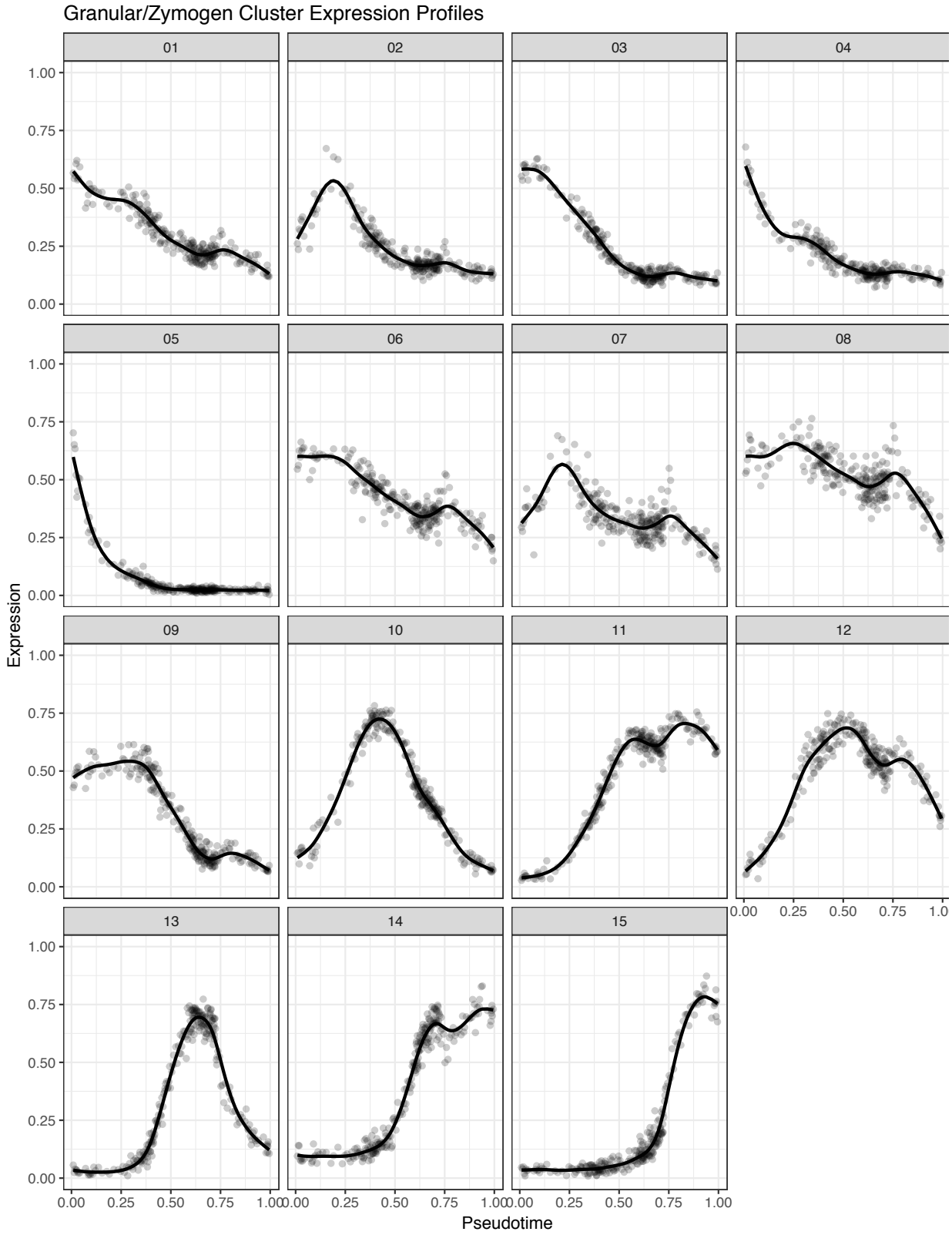
Ectoderm Spatial Expression



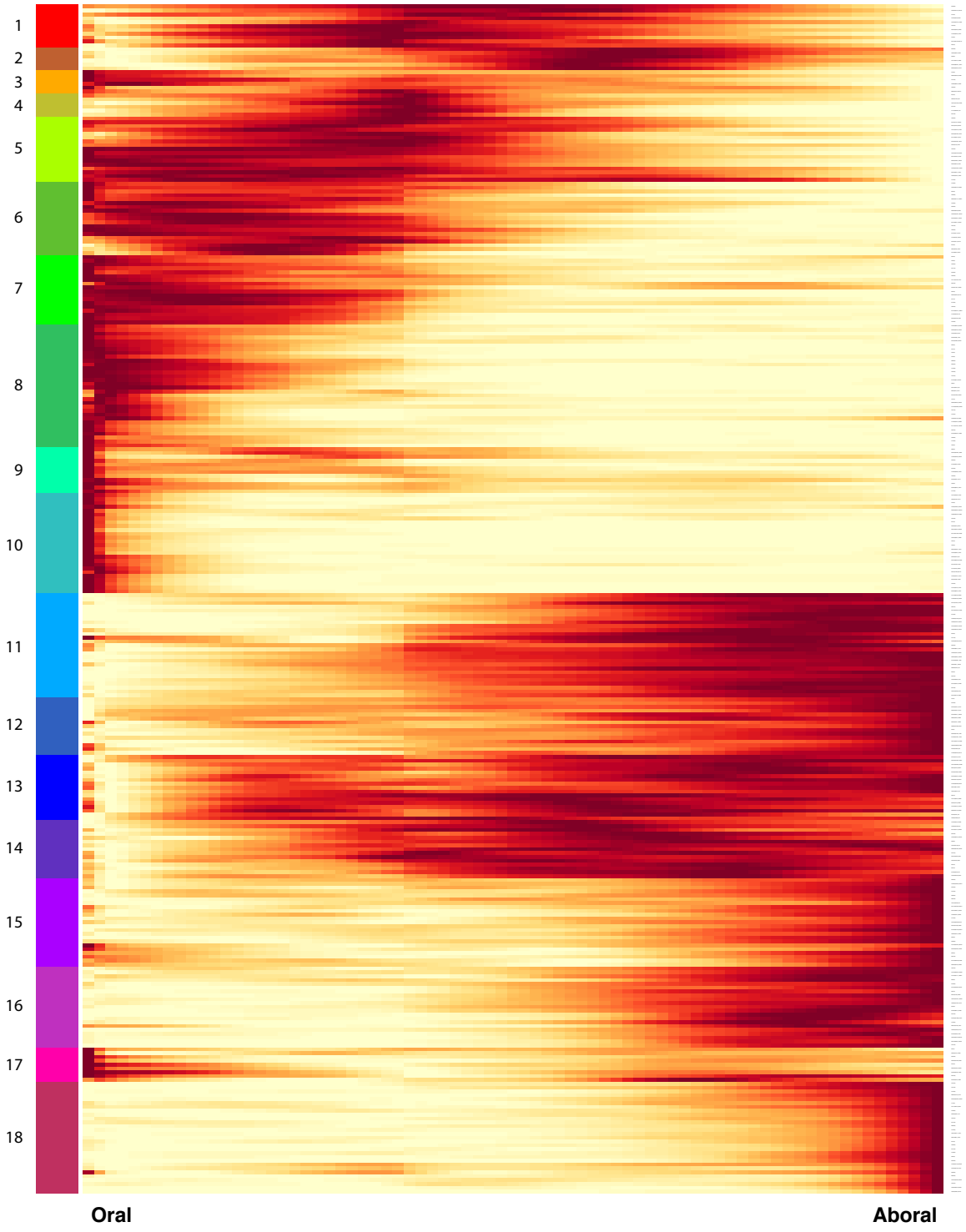

Ectoderm Spatial Cluster Expression Profiles
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
Expression
Pseudotime
0.0
0.4
0.8
0.00
0.25
0.50
0.75
Type
Basal Disc
Body Column
Hypostome
Tentacle

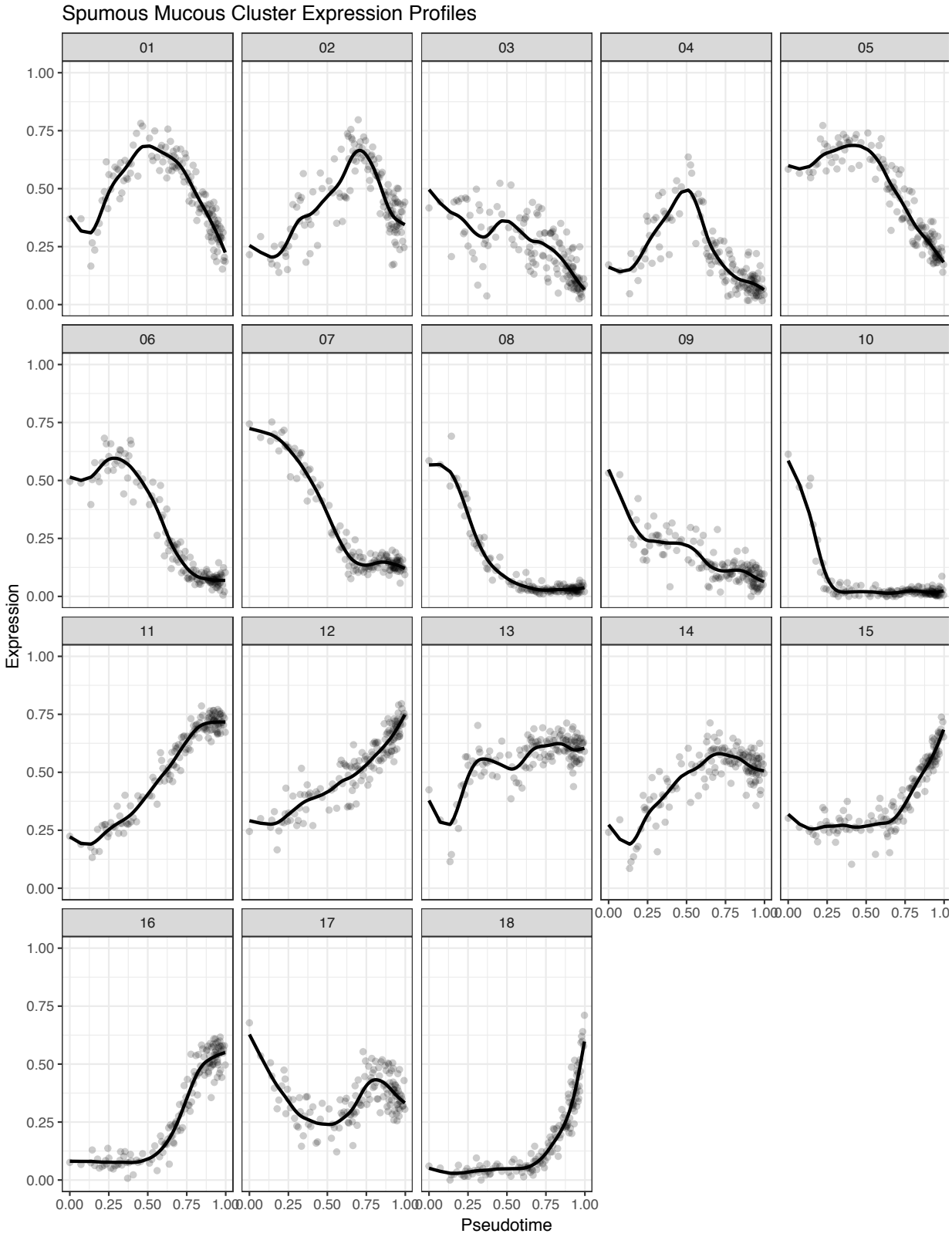
Granular/Zymogen Expression



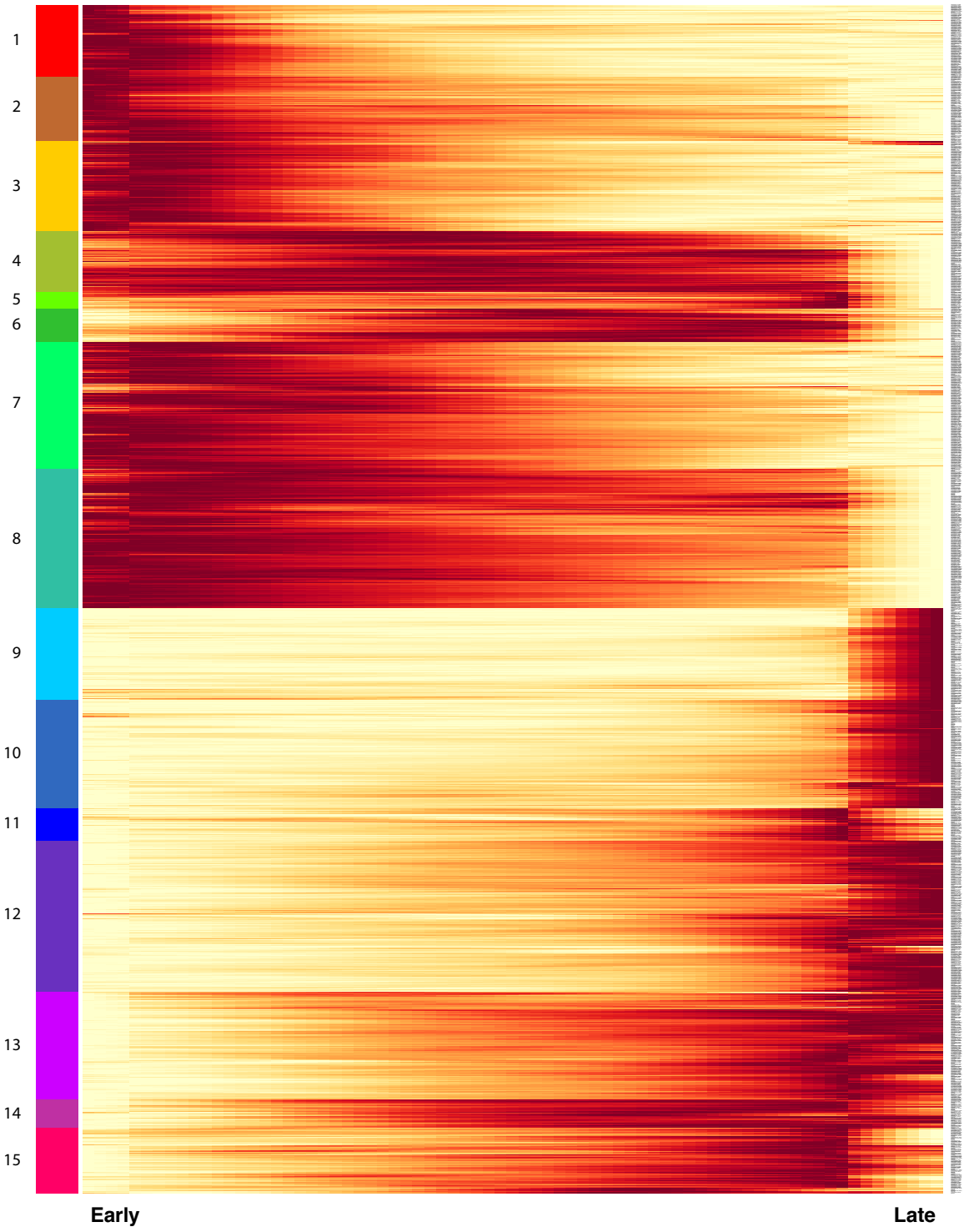

Granular/Zymogen Cluster Expression Profiles
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
Expression
Pseudotime
0.00
0.25
0.50
0.75
1.00

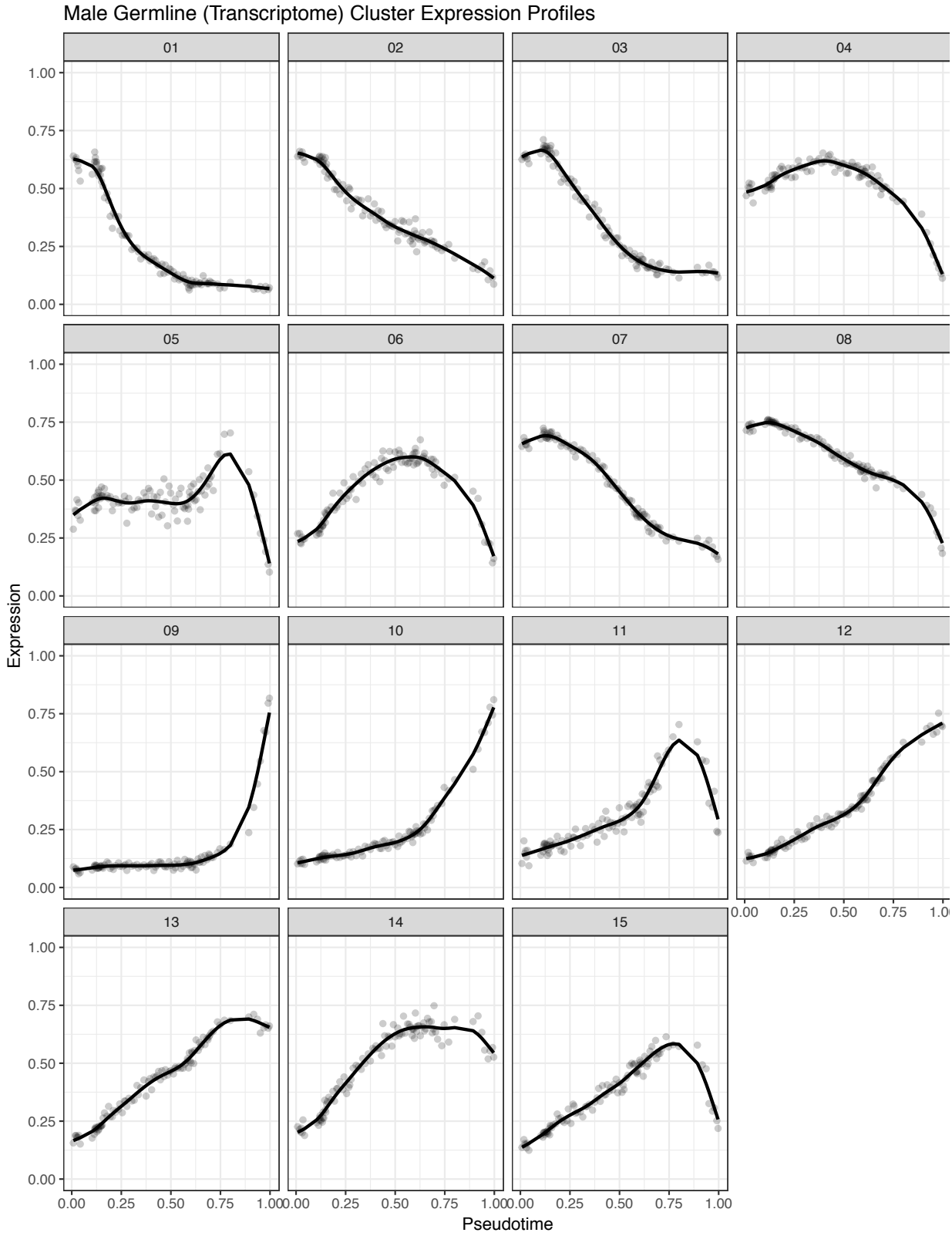
Spumous Mucous Expression



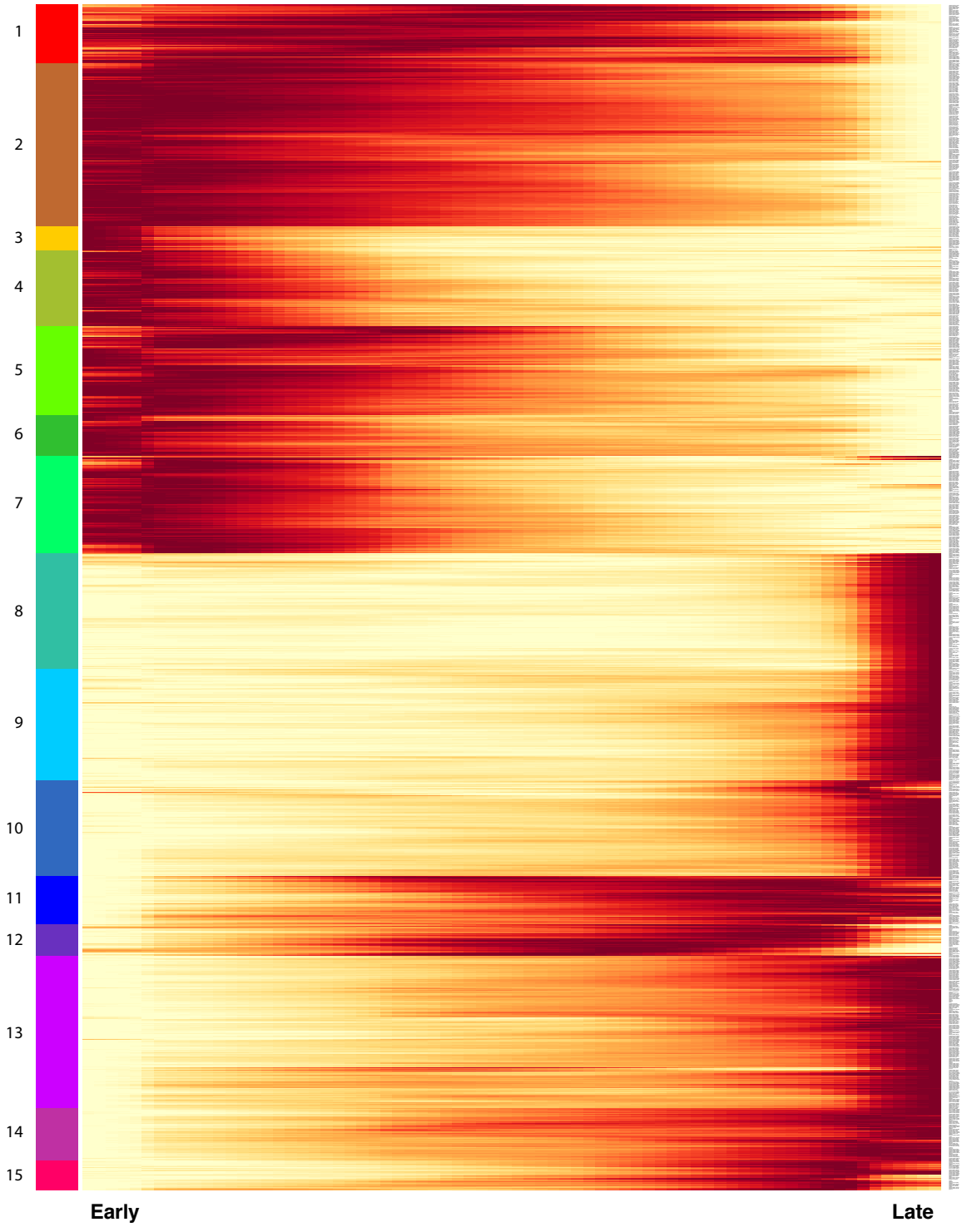

Spumous Mucous Cluster Expression Profiles
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
Expression
Pseudotime
0.00
0.25
0.50
0.75
1.00

Male Germline (Transcriptome) Expression




Male Germline (Transcriptome) Cluster Expression Profiles
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
Expression
Pseudotime
0.00
0.25
0.50
0.75
1.00

Male Germline (Genome) Expression



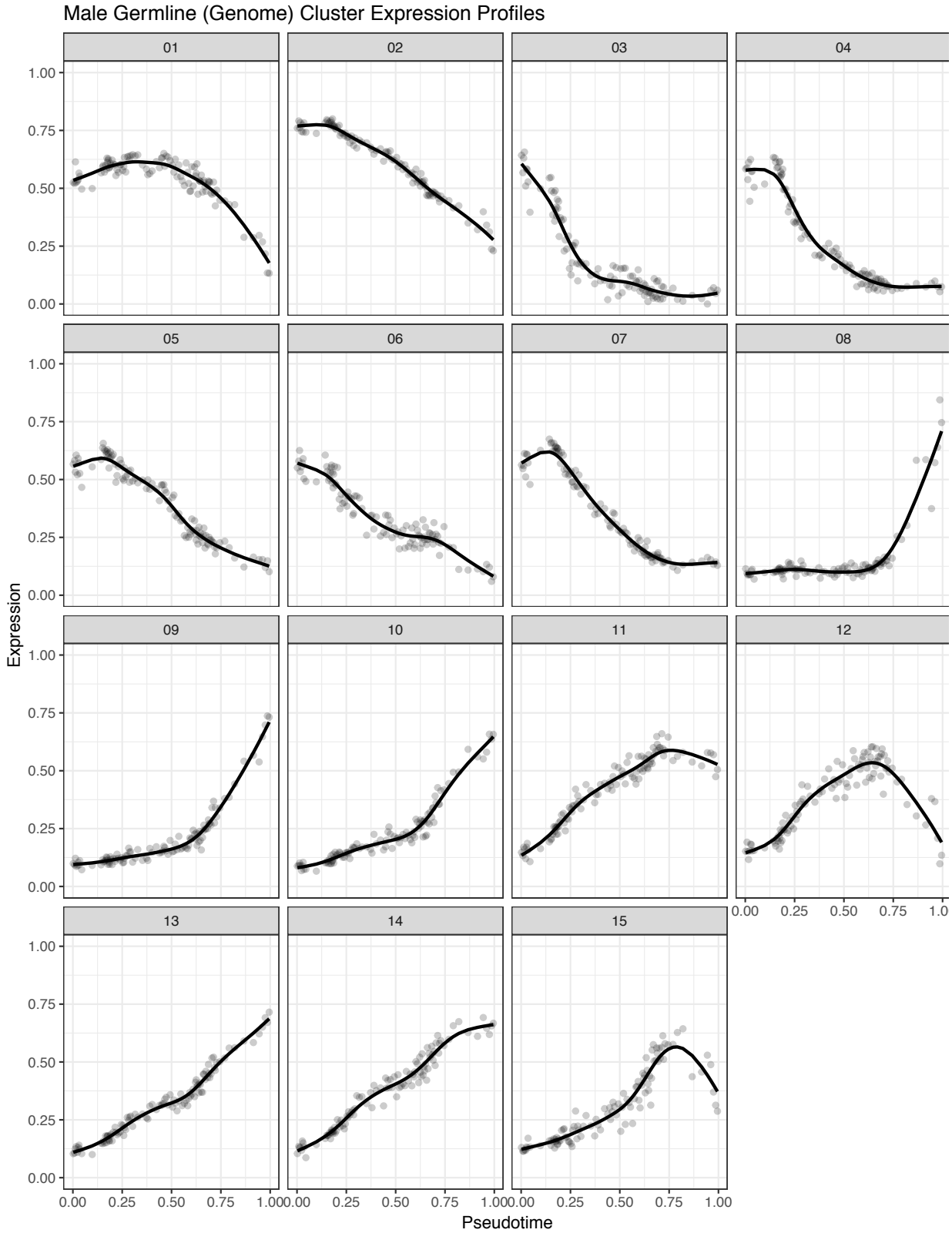


Figure 1.20. Gene expression cascades for epithelial, gland cell, and male germline trajectories. URD trajectories were constructed for endodermal and ectodermal epithelial cells, granular/zymogen gland cells, spumous gland cells, and male germline cells. Germline

trajectories were constructed for transcriptome and genome mapped data. Pseudotime is a proxy for spatial position in case of the epithelial (head/foot) and gland cell trajectories (oral/aboral) and for differentiation progress in case of the male germline (early/late). Genes were selected in each trajectory that varied significantly along pseudotime. Expression is displayed as mean expression of groups of 5 cells, smoothed using a spline curve, scaled to the maximum observed expression (low expression is yellow, high expression is red). Groups of cells are ordered along the x-axis according to pseudotime. In epithelial trajectories, the foot is presented towards the left and the head towards the right. The branched hypostome/tentacle trajectory is represented as two parallel columns of cells; their left sides represent the branchpoint and are concurrent in pseudotime. In all trajectories, genes are ordered along the y-axis by hierarchical clustering to group similar expression profiles, and a color bar along the left side identifies the genes that belong to each cluster. This figure occupies the next 12 pages and comprises a heatmap and the consensus expression profiles for all trajectories visualizing spatially varying genes (for details see Materials and Methods and supplementary analyses SA09-10, SA12-15). Cluster heatmaps from all trajectories can be explored at higher resolution in supplementary file "Gene_expression_cascades".

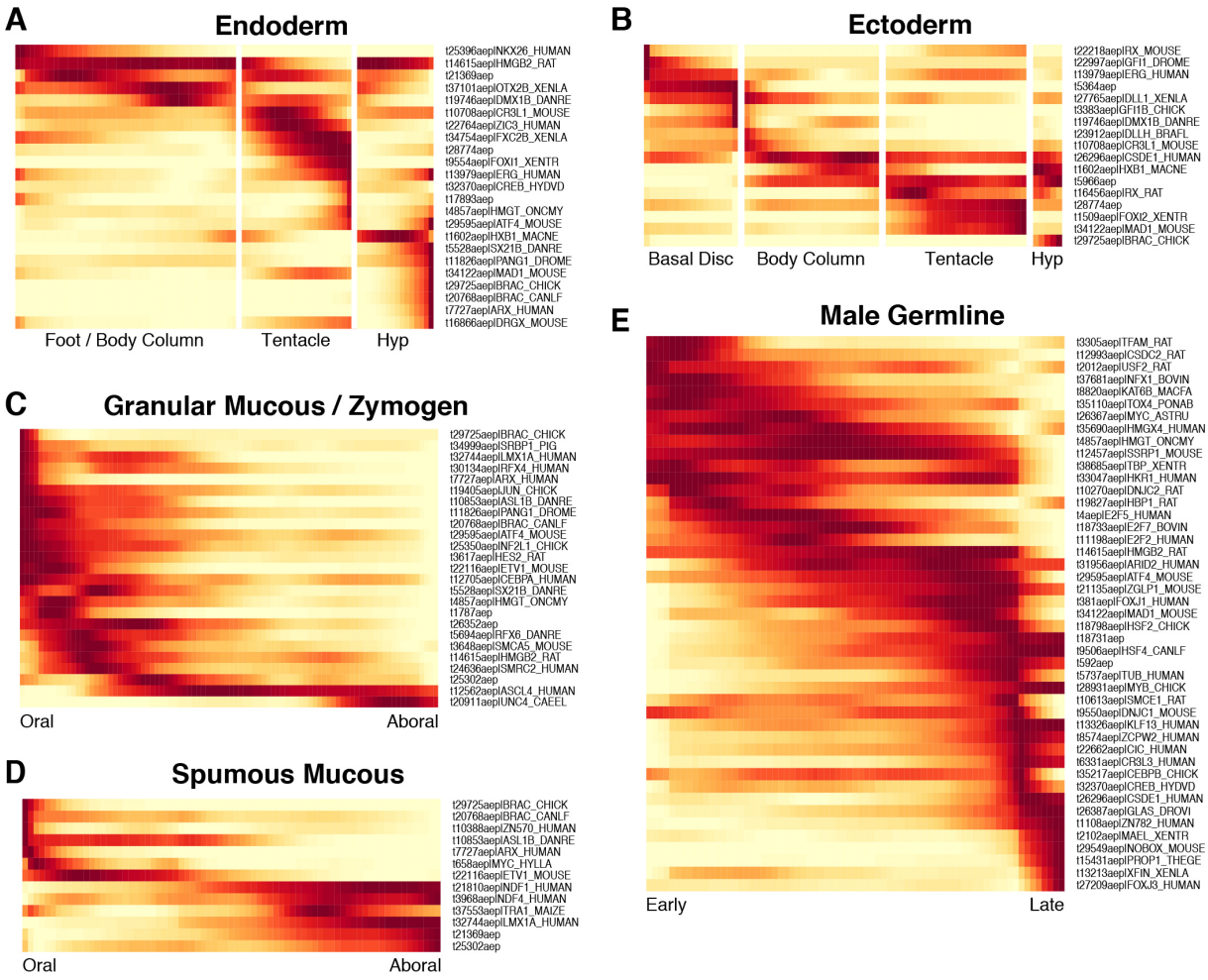


Figure 1.21. Transcription factors with varying expression in branched and linear trajectories for epithelial cells, gland cells and male germline cells. Transcription factors were identified using PFAM annotations. A PFAM annotation file for transcription factors that were identified in the transcriptome reference is available in the accompanying git repository. (A) Transcription factors that vary in endodermal epithelial cells. (B) Transcription factors that vary in ectodermal epithelial cells. RNA in situ hybridization confirming localized expression of *HXB1* (t1602) in both endodermal and ectodermal cells is presented in Figure 1.2 M. (C) Transcription factors that vary in the granular mucous/zymogen (gMGC/ZMG) gland cell population (Figure 1.4 B-C and Figure 1.33). (D) Transcription factors that vary in the spumous mucous gland cell (sMGC) population (Figure 1.4 D and Figure 1.34). (E) Transcription factors that vary in the male germline as identified for transcriptome mapped data (Figure 1.35).

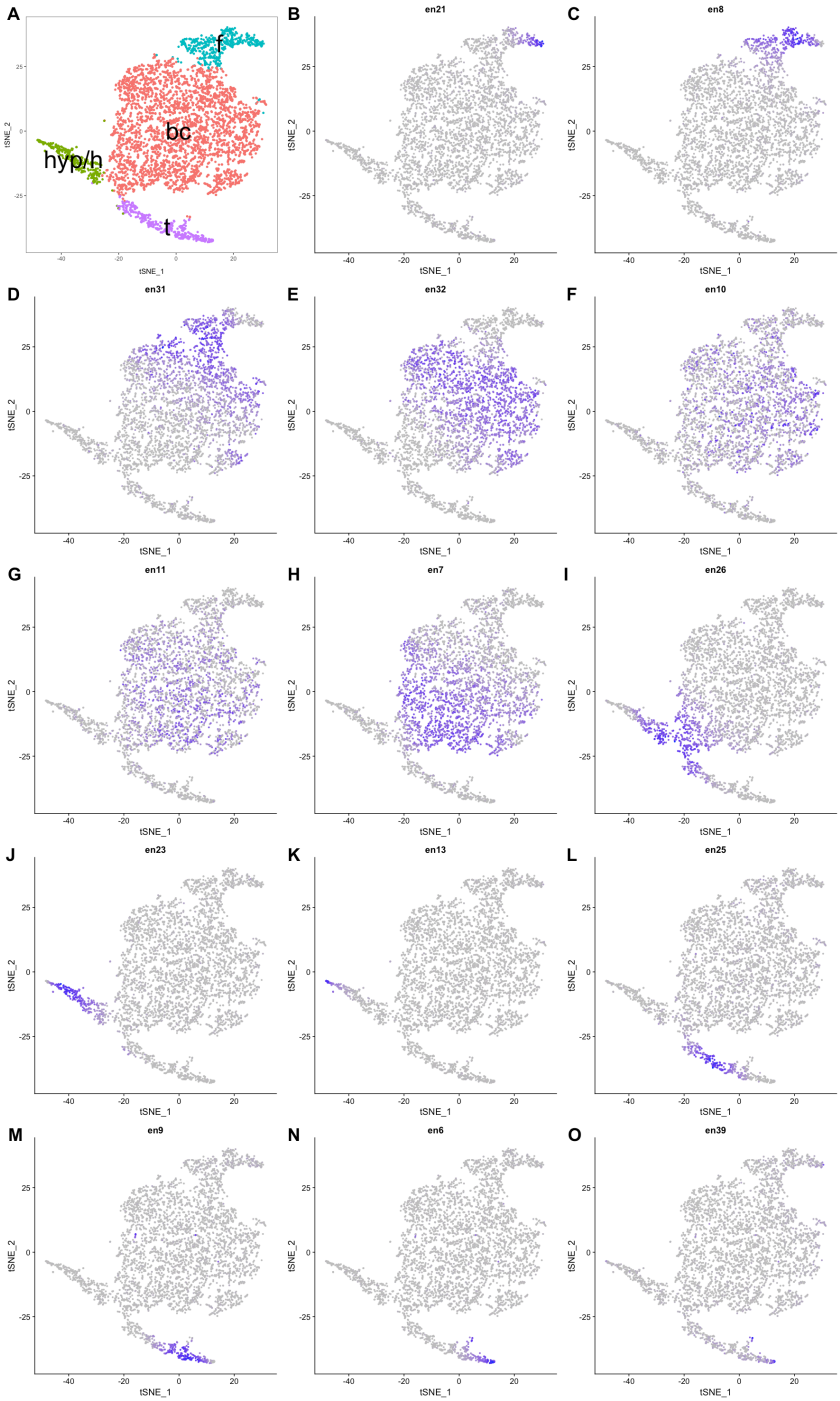


Figure 1.22. Endodermal metagenes (NMF analysis en_K40) expressed along the body column. (A) Annotated t-SNE representation of the endodermal epithelial cell subset, (B-O) Visualized are expression scores for metagenes. Each metagene is composed of a set of co-expressed genes.

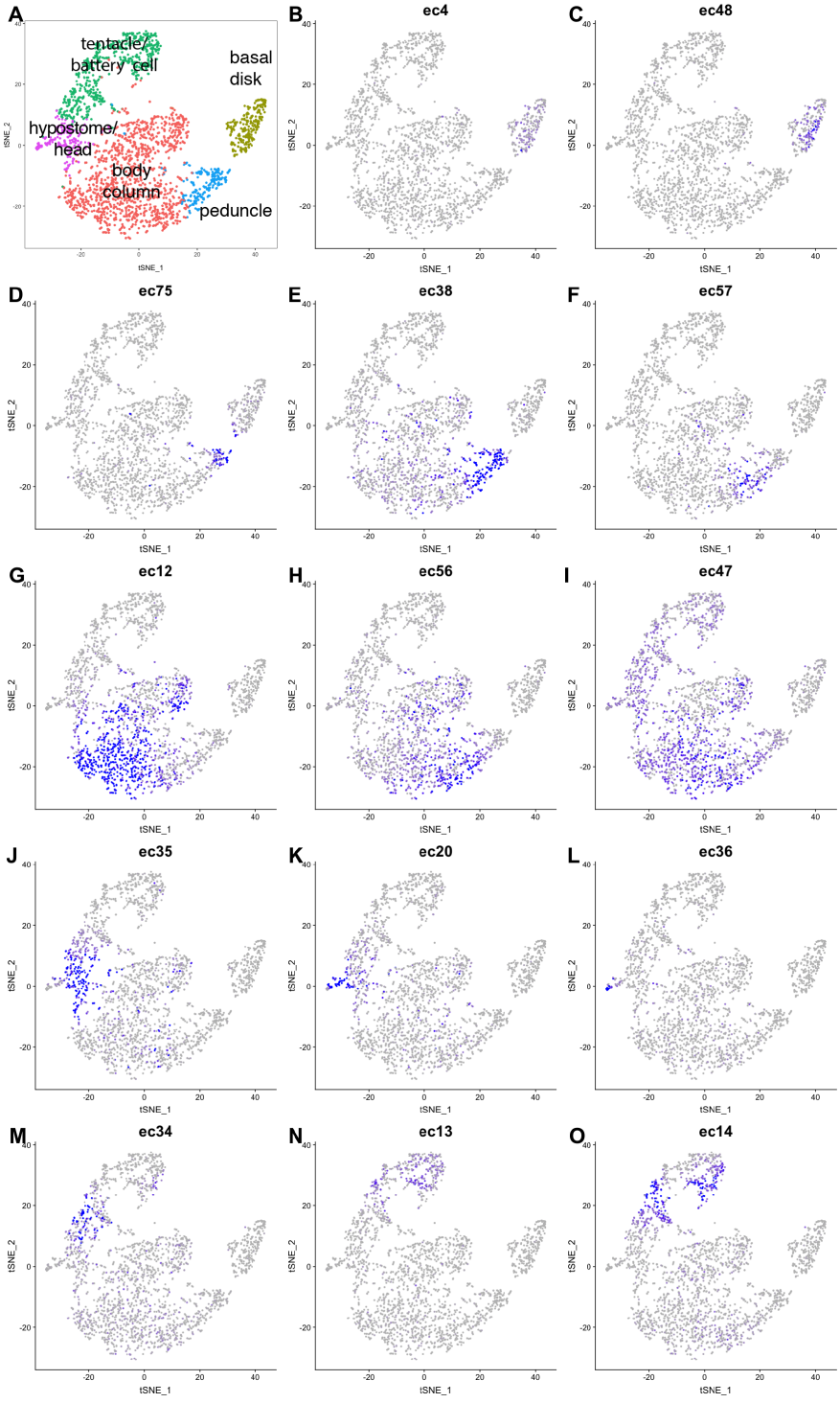


Figure 1.23. Ectodermal metagenes (NMF analysis ec_K76) expressed along the body column. (A) Annotated t-SNE representation of the ectodermal epithelial cell subset, (B-O) Visualized are scores for metagenes. Each metagene is composed of a set of co-expressed genes.

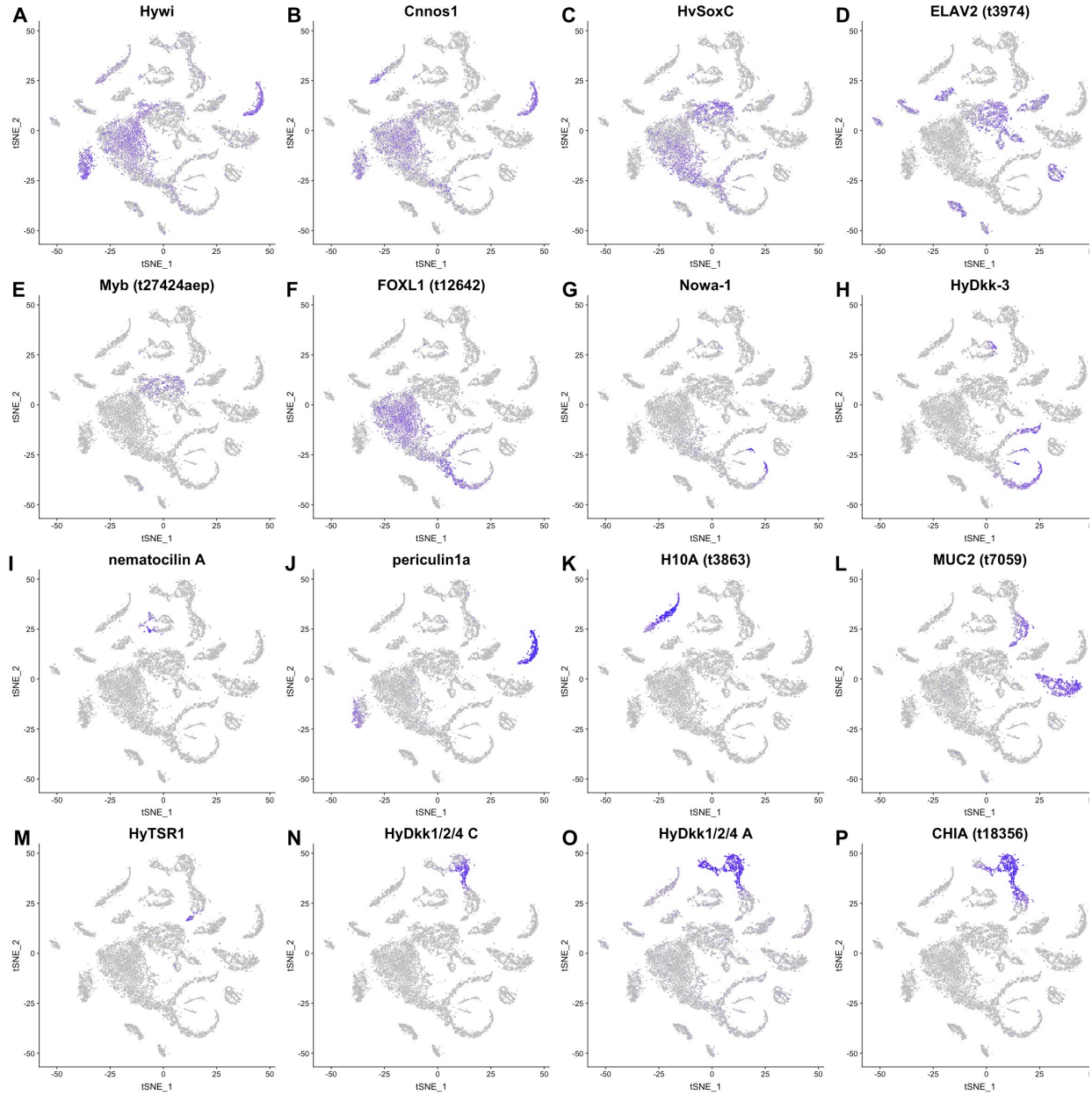


Figure 1.24. Selected markers used for interstitial lineage cluster annotation. (A) Interstitial stem cells, progenitor cells, germline – *hywi* (*122*, *123*). (B) Interstitial stem cells, progenitor, germline - *Cnnos1* (*106*). (C) Differentiating progenitors - *HvSoxC* (this study) (Hemmrich et al., 2012). (D) Neuronal cells - *ELAV2* (t3974) (this study, Figure 1.46 L-M). (E) Neurogenesis, gland cell differentiation - *Myb* (t27424) (this study, Figure 1.27). (F) Interstitial stem cells, nematoblasts - *FOXL1 (t12642)* (this study, Figure 1.26 B). (G) Nematoblasts - *Nowa-1* (Engel et al., 2002). (H) Nematoblasts - *HyDkk-3* (Fedders et al., 2004). (I) Differentiated nematocyte - *nematocillin A*. (J) Female germline - *periculin1a* (Fraune et al., 2010). (K) Male germline - *histone H10A* (t3863) (this study, Figure 1.36 B). (L) Granular and spumous mucous gland cells - *MUC2* (t7059). (M) Granular mucous gland cells - *HyTSR1* (*15*). (N) Zymogen gland cells - *Hydkk1/2/4 C* (Augustin et al., 2006; Guder et al., 2006). (O) Zymogen gland cells - *Hydkk1/2/4 A* (Augustin et al., 2006; Guder et al., 2006). (P) Zymogen gland cells - *CHIA* (t18356) (this study, Figure 1.32 B-C).

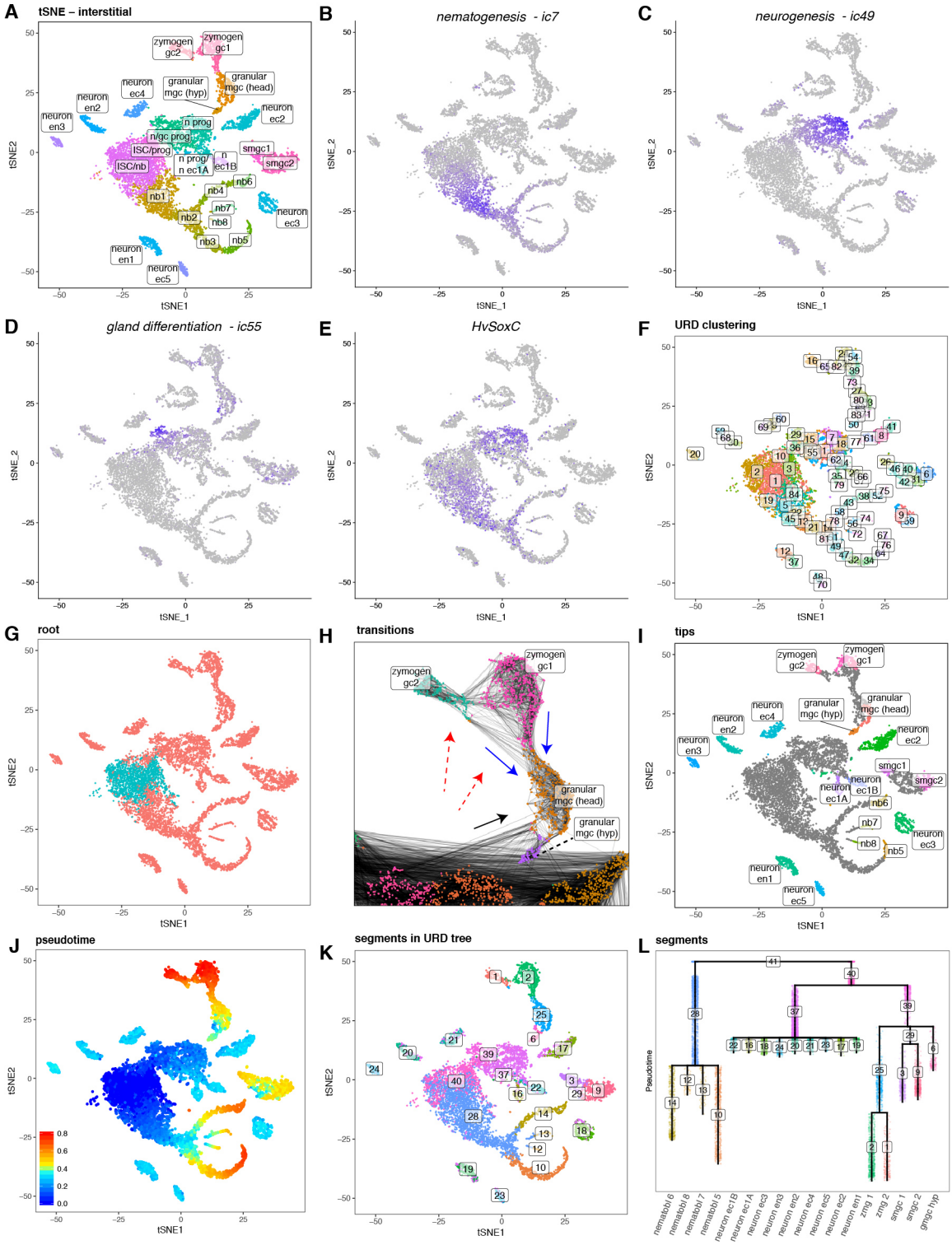


Figure 1.25. URD reconstruction of differentiation trajectories for cells of the interstitial cell lineage. Germ cells were excluded from this analysis because germline stem cell (GSC) differentiation from ISCs does not typically occur in a homeostatic animal (Nishimiya-Fujisawa

and Kobayashi, 2012). (A) Annotated t-SNE representation after exclusion of germline cell clusters and the cluster containing mature nematocytes. Cluster "nematocyte" (Figure 1.3 (A) was found to contain fully differentiated nematocytes based on expression of known mature differentiated nematocyte markers (Figure 1.30 A-C), but we could not unambiguously assign four expected terminal fates. This cluster was therefore excluded from the trajectory analysis. gc: gland cell, hyp: hypostome, ISC: interstitial stem cell, mgc: mucous gland cell, n: neuron, nb: nematoblast, prog: progenitor, smgc: spumous mucous gland cell. (B-D) Differentiation gene modules identified using NMF analysis. Selected top scoring genes in each metagene were validated via RNA in situ hybridization. (B) Interstitial gene module ic7 is expressed during nematogenesis (e.g. gene *TEAD* (t33926), Figure 1.31). (C) Interstitial gene module ic49 is expressed during neurogenesis (e.g. *ELAV2* (t3974), Figure 1.46 L-M). (D) Interstitial gene module ic55 is expressed during gland cell differentiation (e.g. *COMA* (t2163), Figure 1.27 E). (E) *HvSoxC* was identified as a putative general differentiation marker that is expressed in cells undergoing nematogenesis, neurogenesis, and early gland cell differentiation (compare to B-D). (F-G) Three cell clusters (clusters 1, 2 and 3 from separate clustering within URD) were chosen as the root (starting point) for URD trajectory reconstruction. Cells in this population (G, turquoise) reside in between cells that are in the process of differentiation as indicated by metagene expression (see B-D), are proliferating as indicated by *PCNA* expression (Figure 1.30 J), and express genes like piwi and nanos that were previously shown to be expressed in ISCs (Figure 1.24 A-B). These features are consistent with the interpretation that these cells are multipotent interstitial stem cells. (H) Choosing terminal gland cell fates for trajectory reconstruction: URD trajectory reconstruction involves calculating transition probabilities between cell states and the transition probabilities between cells are visualized on the t-SNE representation as black lines. Transitions were identified from progenitors to head granular mucous gland cells (black arrow) but no direct transitions were observed between progenitors and zymogen gland cells (red arrows). This may indicate that homeostatic animals have a low number of zymogen gland cell progenitors and that zymogen gland cell production may primarily occur through mitotic division of existing zymogen gland cells. As they move along the body column, zymogen gland cells can transition into head gland cells, and this transition of states is reflected in the data (blue arrows). We therefore chose to include zymogen states as trajectory tips in the reconstruction. Thus, in the resulting trajectory, zymogen states are connected to progenitor states via granular mucous gland cells in the *Hydra* head (cells in segment 25 (K-L)). (I) Terminal cell populations that were selected as endpoints for the trajectory reconstruction. (J) Cells were ordered based on their expression and pseudotimes were assigned. A strong pseudotime gradient is observed in nematogenesis and granular mucous gland cells/zymogen gland cells and a moderate pseudotime gradient is observed in the spumous mucous gland cells and neurons. Not all terminal populations have the latest pseudotime because this data is not a time-course, but a profile of a homeostatic animal. Therefore, the length of each trajectory in pseudotime is reflective of the transcriptional differences it exhibits compared to the stem cell population. (K) Segments (numbered) from URD tree visualized on the t-SNE representation. (L) URD differentiation tree resolving nematogenesis, neurogenesis, and gland cell differentiation trajectories. The recovered segments are numbered.

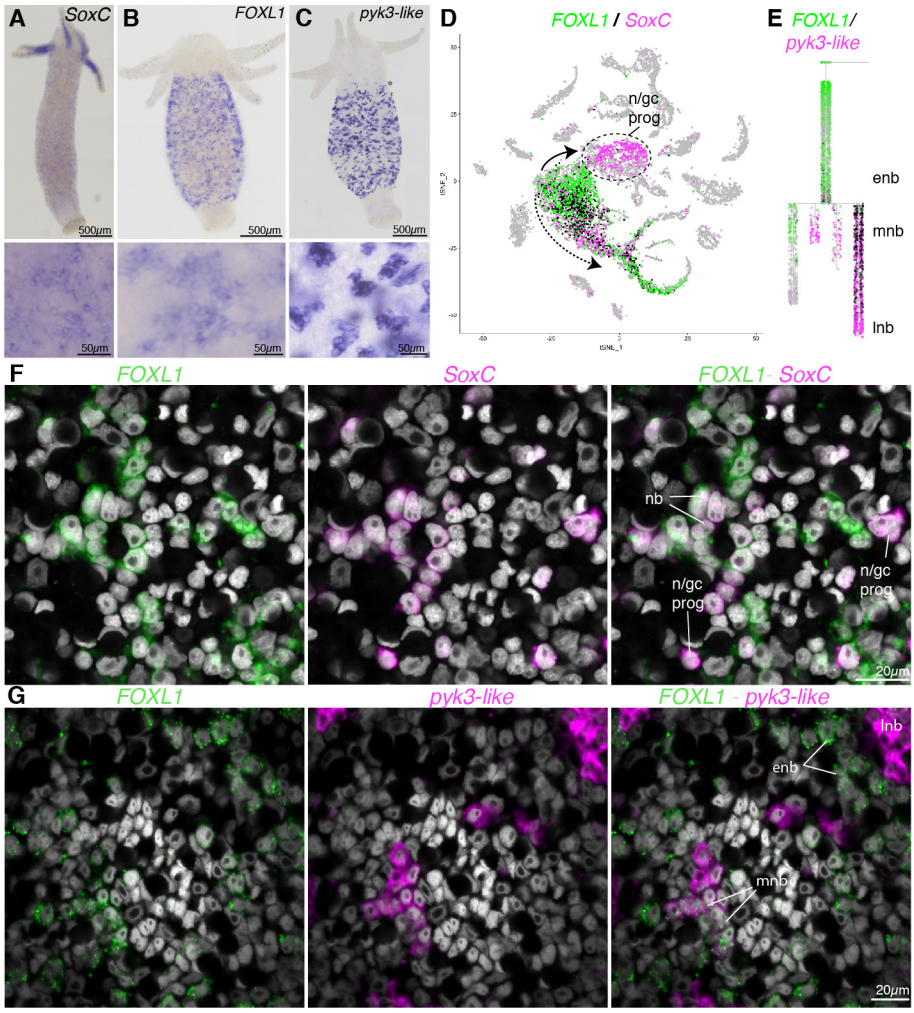


Figure 1.26. Validation of cell states. (A-C) Colorimetric RNA in situ hybridization for selected interstitial genes. Whole mounts and mid-body close-ups for *HvSoxC* (A), *FOXL1 (t12642)* (B), and *pyk3-like* (t15240) (C). (D) Gene expression visualized on the interstitial cell clustering for *HvSoxC,* which is expressed in neuronal/gland progenitors and nematoblasts, and *FOXL1,* which is expressed in putative ISCs and throughout nematogenesis. scRNA-seq data reveal partial overlap in the expression domains of *FOXL1* and *HySoxC*; the co-expression of *HySoxC* and *FOXL1* is found in the nematoblast cell states. Co- expressing cells are black. Solid arrow: neurogenesis/gland cell differentiation. Dashed arrow: nematogenesis. (E) Gene expression visualized on the URD trajectory tree (nematogenesis branch) for *FOXL1 and pyk3-like* (t15240). scRNA-seq data reveal partial overlap in the expression domains of *FOXL1* and *pyk3-like*; the co-expression of *FOXL1* and *pyk3-like* occurs in mid-stage nematoblasts. Co-expressing cells are black. (F-G) Double fluorescent RNA in situ hybridization results are consistent with predictions. Double FISH for *FOXL1* and *HvSoxC* (F) and *FOXL1 - pyk3-like* (G). In both cases, we detect co-expressing cells and cells that express the markers individually. Co-expression in clusters of cells is indicative for nematoblast nests. Cells expressing only *FOXL1* in (F) include multipotent stem cells, early progenitors, and late nematoblasts. enb: early nematoblast, lnb: late stage nematoblast, mnb: mid-stage nematoblast, nb: nematoblast, n/gc prog: putative neuronal or gland cell progenitors.

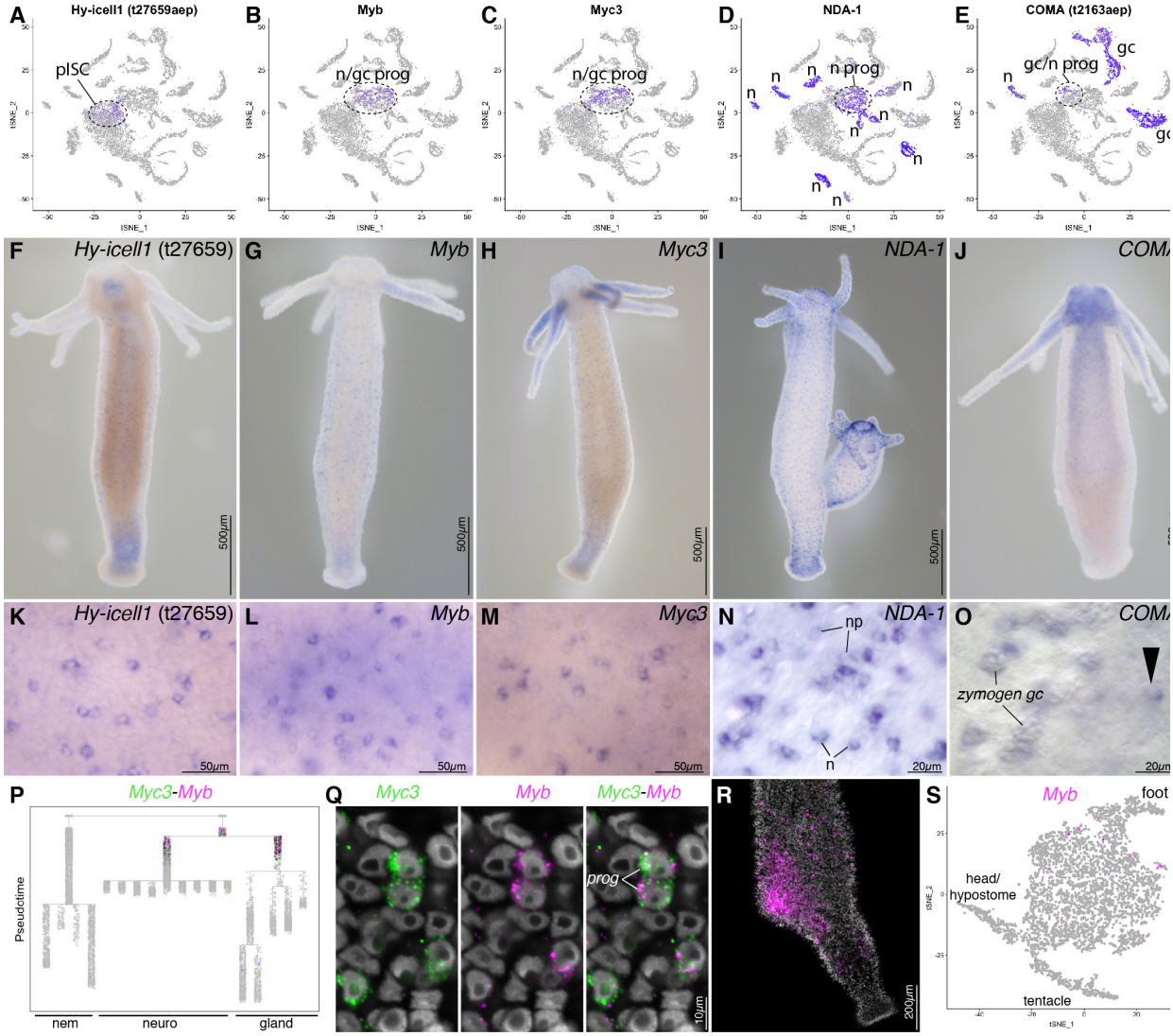


Figure 1.27. Biomarkers for interstitial cell populations. (A-O) Expression plots, RNA in situ hybridization: whole mounts and close-ups. (A,F,K) Putative interstitial stem cell (pISC) marker *Hy-icell1* (t27659). (B,G,L and C,H,M) *Myb* (t27424) and *Myc3* (t18095) expressed in the shared neuro/gland progenitor cell state and during early neurogenesis and gland cell formation. (D,I,N) *NDA-1* (Augustin et al., 2017) was used as a marker for differentiated neurons and late neurogenesis. (E,J,O) *COMA* (t2163) was used as a marker for differentiated gland cells and gland cell differentiation. *COMA* (t2163) is expressed in a subset of neurons. (K-M) Mid–body-column close-ups. (N) Region in between tentacles revealing strongly stained differentiated neurons (n) and lighter stained cells indicative of differentiating neuronal progenitors (np) (see Figure 1.28). (O) Upper body column close-up, endodermal layer, with expression in zymogen gland cells and unidentified cells (arrow). (P) *Myb* and *Myc3* gene expression visualized on the URD trajectory tree reveals largely overlapping expression domains. Co-expressing cells are black. (Q) Double fluorescent RNA in situ hybridization results confirm cell states that co-express both genes. (R) Fluorescent RNA in situ hybridization revealed *Myb* expression in endodermal epithelial cells of developing buds. (S) *Myb* expression

visualized on the endodermal epithelial cell subclustering reveals enrichment of positive cells in suspected budding zone cells adjacent to the foot cluster. We collected 14 Drop-seq libraries using *Hydra* polyps that did not bear apparent buds to collect the cell complement of a homeostatic *Hydra* (Table 1.2). Only library 11-BU intentionally included budding polyps (Table 1.1). This library contributed 3,207 cells or 12.8% to the full data set (Table 1.3). 60.7% (17/28) of *Myb* positive endodermal epithelial cells with normalized expression > 0.5 originated from this library.

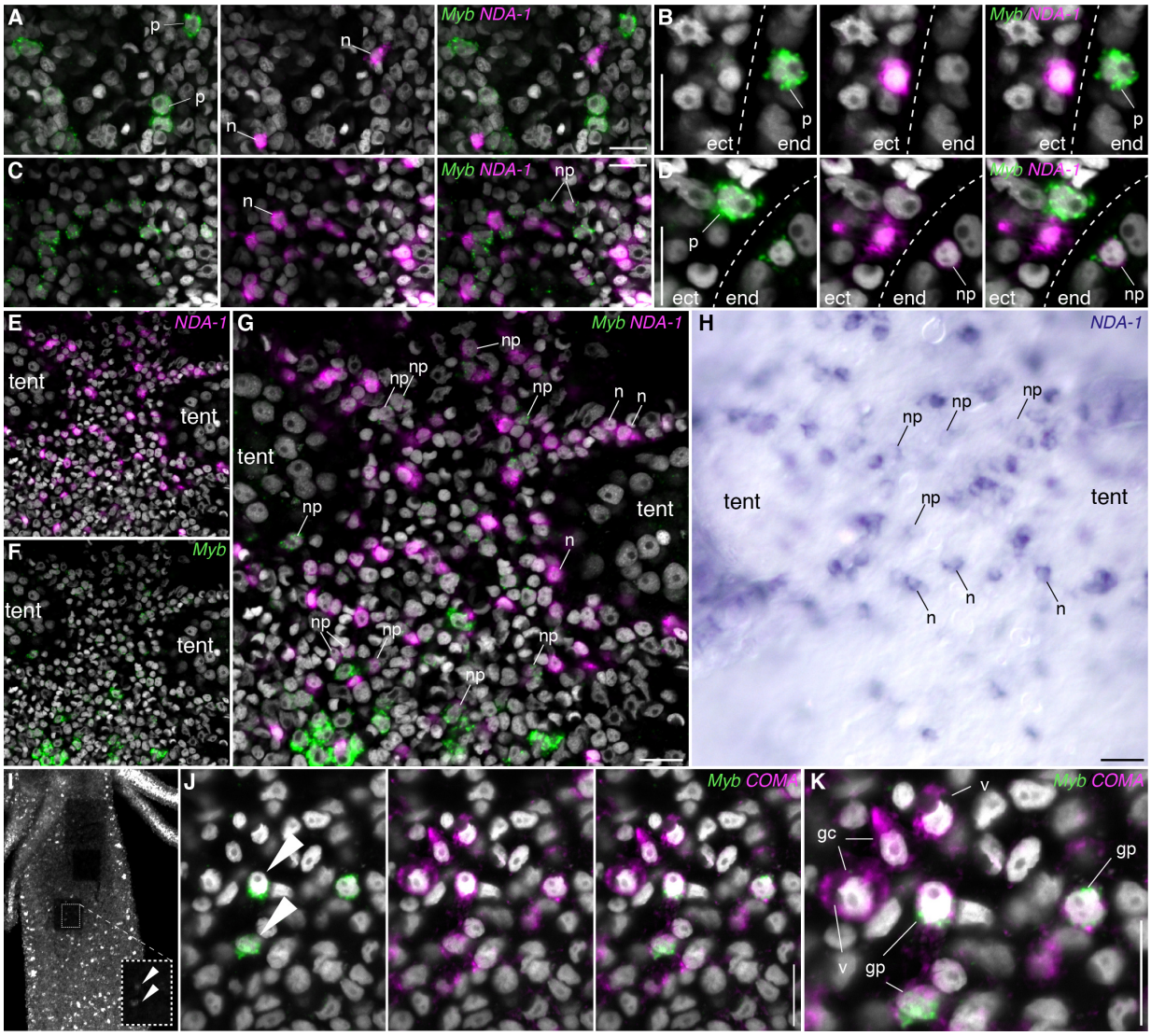


Figure 1.28. Validation of predicted progenitor transition states using double fluorescent RNA in situ hybridization (FISH). In A-G and J both the separate channels and the overlay are shown. (A-G) Co- expression analysis for *Myb* (t27424) (expressed in early progenitors) and *NDA-1* (strongly expressed during late neurogenesis and in differentiated neurons) reveals neuroblasts in ectoderm and endoderm. (A) Ectoderm (mid-way between tentacles and budding zone) with *Myb* positive progenitors (green) and *NDA- 1* positive differentiated neurons (magenta). (B) Body column cross-section (ring) of the subtentacle region revealing a *Myb* positive progenitor in the endoderm. (C) Head of developing bud (already bearing tentacle buds), a site of active neurogenesis, revealing ectodermal neurons in a transition state expressing both *Myb* and *NDA-1*. Note the smaller cell size of the differentiating cells as compared to the progenitors shown in (A). (D) Upper body column cross-section (ring) revealing endodermal neuron progenitors expressing both *Myb* and *NDA-1*. (E-G) We anticipated the presence of progenitors migrating from the body to the intertentacle region to supply the neuron-dense head and hypostome with neurons. Double FISH reveals these expected progenitors. (H) Colorimetric RNA in situ hybridization for *NDA-1* of similar area that is shown in (G) revealing differentiated neurons (strong *NDA-1* expression) and differentiating progenitors (weak *NDA-1* expression). (I) Region of analysis for double FISH using probes for *Myb* (green) and *COMA* (t2163,

magenta) co-expression. *COMA* has expected expression in gland cell differentiation, differentiated gland cells and a subset of endodermal neurons. *Myb* signal of two cells shown in (J) is observed in the boxed region (arrows). (K) Enlargement of region in (J). (J,K) Large cells positive for both *Myb* and *COMA* transcripts can be observed next to cells with gland cell morphology characterized by the presence of secretory vesicles. Gland cells and progenitors are of comparable size. This is consistent with gland cell differentiation occurring in the absence of cell proliferation (Bode, 1996). By contrast, cell divisions during neurogenesis produces cells smaller than the progenitors (compare *Myb* positive progenitors versus neuronal progenitors (A versus C, D) (Bode, 1996). p: early progenitor, np: neuronal progenitor, gp: gland progenitor, v: vesicles, tent: tentacles, n: neuron. All scale bars: 20 μm.

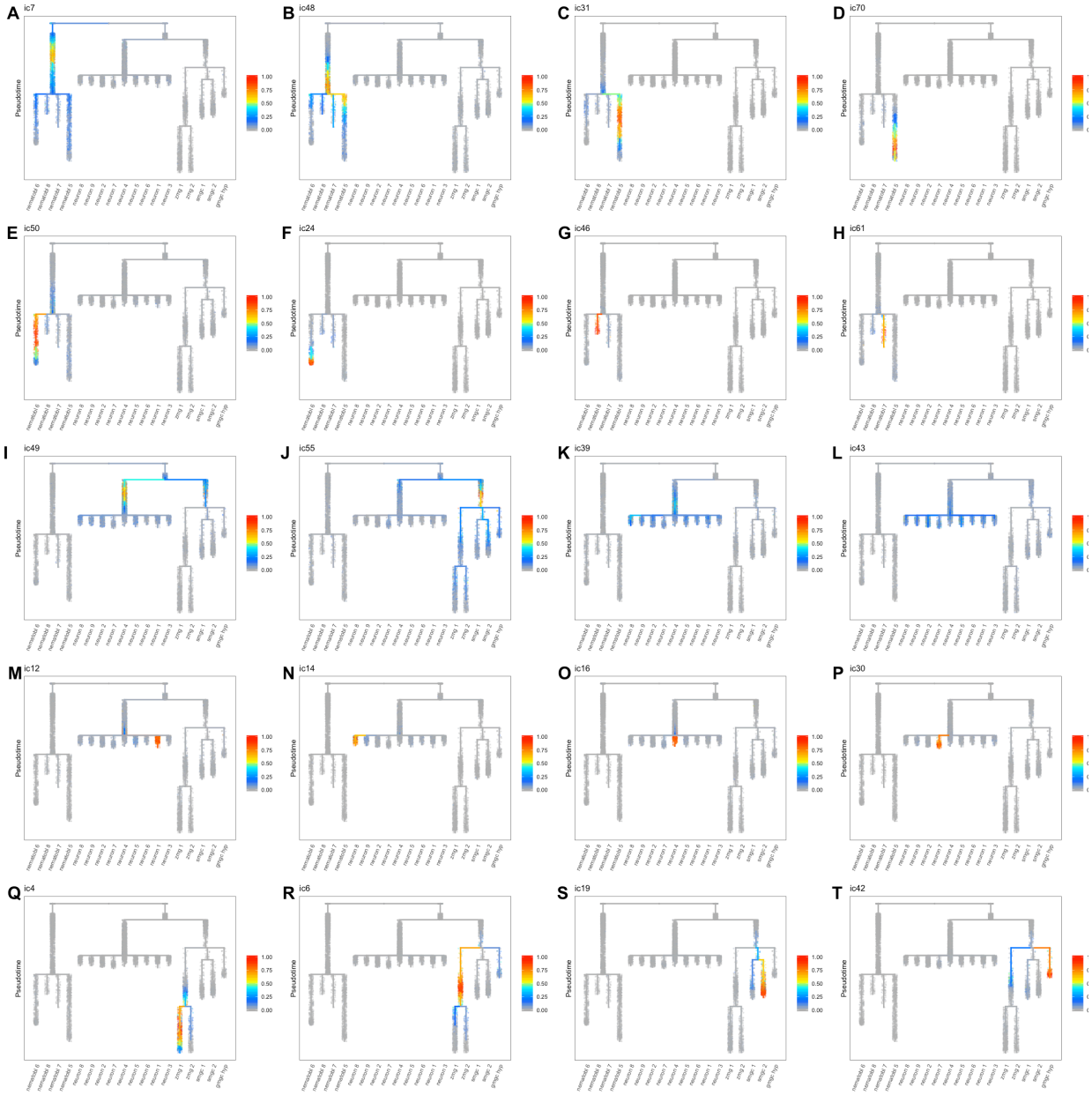


Figure 1.29. **Gene modules expressed along interstitial cell differentiation trajectories.** NMF analysis (ic_K75) identified gene modules with ordered expression along trajectories. (A-H) Gene modules expressed during nematogenesis from early to late (A-D) and modules specific to the four recovered branches (D-H). (I,J) Modules expressed during neurogenesis and gland cell differentiation. (K,L) Neurogenesis specific modules. (M-P) Modules expressed in specific neuron subtypes. (Q-T) Modules expressed in gland cell subtypes. (Q) Zymogen gland cells (ZMG) in the upper body column. R) granular mucous gland cells (gMGC) in the *Hydra* head. (S) spumous mucous gland cells (sMGC) in the lower head (smgc2 population, see Figure 1.28 I-K). (T) granular mucous gland cells (gMGC) in the hypostome.

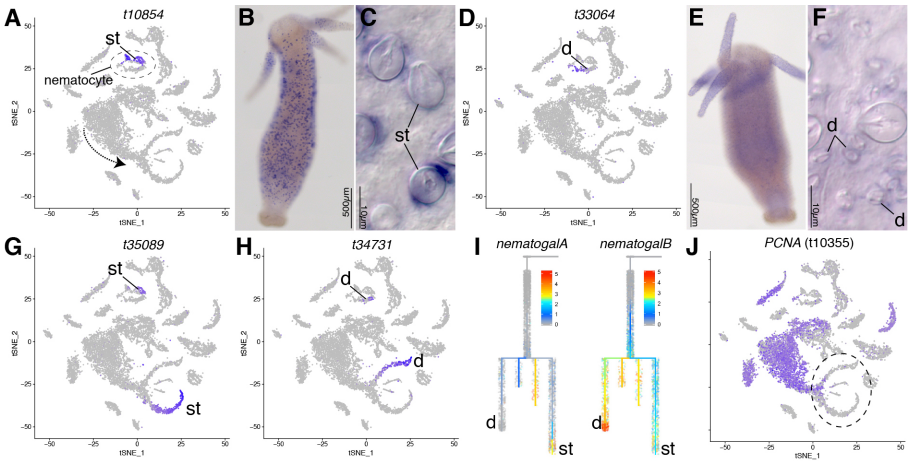


Figure 1.30. Identification of trajectories for nematocytes forming stenoteles or desmonemes. (A-F) Expression plots and RNA in situ hybridizations for transcripts *t10854* and *t33064* highlights as differentiated nematocytes. Arrow in A indicates nematogenesis. (A-C) *t10854* is expressed in stenotele- forming nematocytes (st). (D-F) *t33064* is expressed in desmoneme-forming nematocytes (d). (G) *t35089* expression links a nematoblast trajectory branch to stenotele fate. (H) *t34731* expression links a nematoblast trajectory branch to desmoneme fate. (I) Lack of nematogalectin A expression supports desmoneme (d) nematocyst type assignment. Nematogalectin B is expressed in all trajectories (Hwang et al., 2010). Unannotated branches may correspond to nematocytes containing two types of isorhiza but await further characterization. (J) *PCNA* (t10355) expression marks proliferative progenitor cells and highlights cells in late nematoblast clusters (circled) as postmitotic.

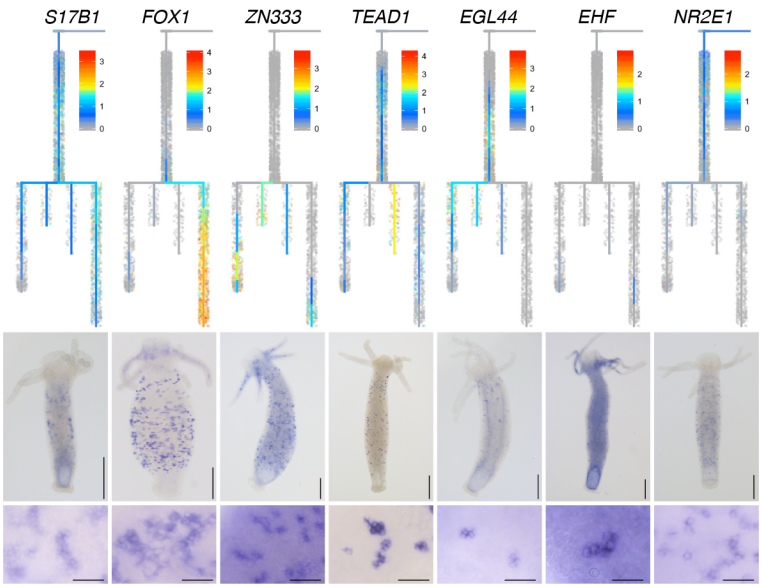


Figure 1.31. Previously uncharacterized transcription factors with differential expression in nematocyte trajectories. Expression was visualized on the URD trajectory tree for transcription factors *S17B1* (t17610), *FOX1* (t16997), *ZN333* (t11591), *TEAD1* (t33926), *EGL44* (t7356), *EHF* (t27653), and *NR2E1* (t28441). Expression was validated using mRNA in situ hybridization, which revealed characteristic expression in nests of cells typical of developing nematoblasts. This demonstrated that our trajectories contain only nematoblasts because these cells undergo incomplete cytokinesis and only resolve into single nematocytes once they mature (Fujisawa and David, 1981) (Figure 1.1 D). The fact that we recovered these cells in the Drop-seq libraries suggested that the fine cytoplasmic bridges connecting the cells rupture during the dissociation procedure but that cells largely remain their integrity. Close-ups show mid-body region. Whole mount scale bars: 500 μm. Close-up scale bars: 50 μm.

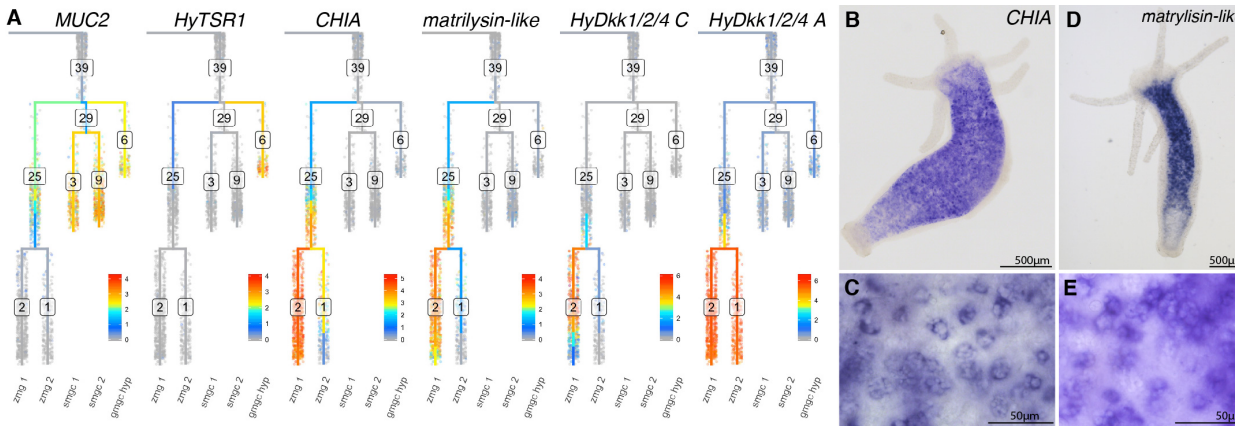


Figure 1.32. Evaluation of the gland cell trajectories recovered in the URD differentiation tree. (A) Gland cell gene expression plotted on the URD differentiation tree. We interpret cells in segments 3,6, and 9 as mucous gland cells based on expression of *MUC2* (t7059). Cells in segment 6 are positive for *HyTSR1,* which is expressed in granular mucous gland cells (gMGC) of the upper head (hypostome). We interpret the remaining mucous positive cells in segments 3 and 9 as the spumous mucous gland cell population (sMGC) in the head. Segment 25 contains *MUC2* - positive gMGCs of the lower head and these form a continuum of states with *MUC2*-negative zymogen gland cells (ZMGs) of the body column (segments 2 and 1). This reflects the position dependent differentiation of ZMGs into gMGCs as they are displaced into the head. The ZMGs in segment 2 show positive signal for several genes expressed in ZMGs throughout the body column, including a gene encoding a chitinase (CHIA, t18356) (B,C), matrilysin-like (t32151)(D,E), *HyDkk1/2/4 C* (*51*, *52*), and *HyDkk1/2/4 A* (*51*, *52*). Segment 1 contains ZMGs predominantly found in the lower body column since they are negative for *HyDkk1/2/4 C* expression but positive for *HyDkk1/2/4 A* expression. This may indicate that this ZMGs state can directly transdifferentiate into head gMGCs without passing through gene expression states present in segment 2; see additional analysis of the ZMG/gMGC gland cell trajectory (Figure 1.33 B-C) RNA in situ hybridization for a chitinase (*CHIA*, t18356), (C) Enlargement of mid-body region reveals zymogen gland cell expression. (D,E) RNA in situ hybridization for *matrylisin-like* (t32151). (C,E) Enlargement of mid-body region reveals zymogen gland cell expression.

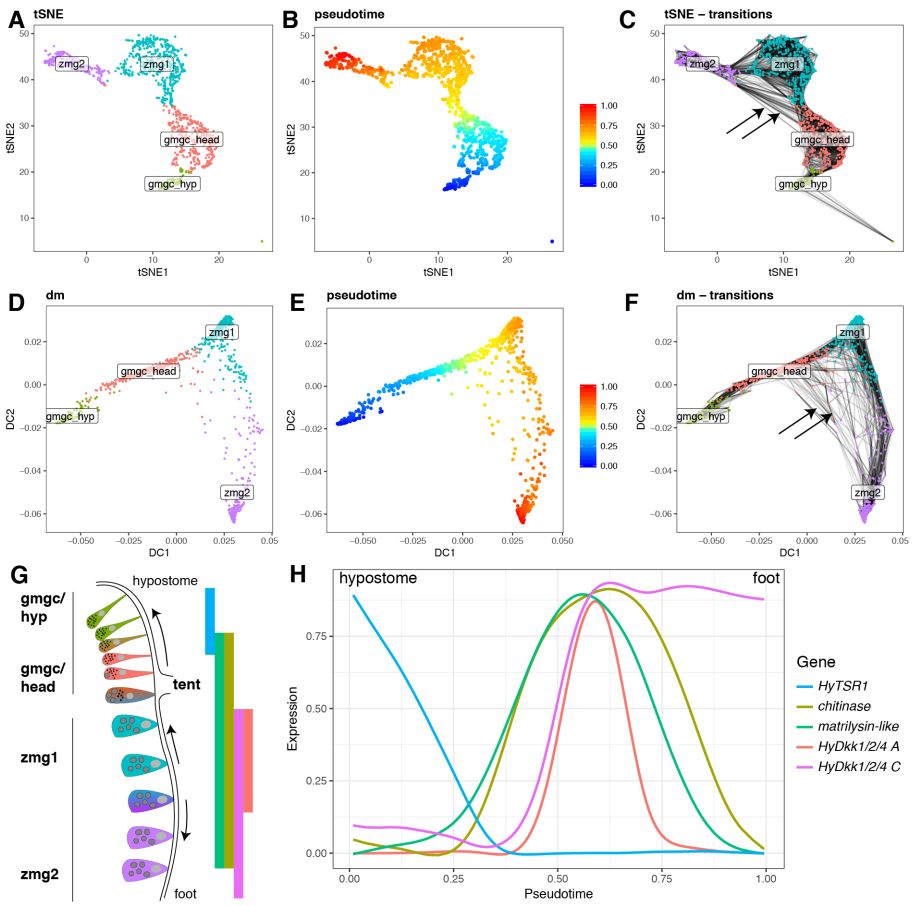


Figure 1.33. URD reconstruction of a linear trajectory for gMGC/ZMG gland cells. (A) t-SNE representation of the gMGC/ZMG cell populations. (B) Cells were ordered based on their expression using hypostomal cells as the root (dark blue) and pseudotimes assigned. (C) t-SNE plot with transition probabilities represented as lines in between cells. (D) URD diffusion map for gMGC/ZMG cell populations. (E) URD diffusion map with pseudotime assigned. (F) URD diffusion map with transition probabilities. Transition probabilities represented as lines in between cells indicate that zmg2 cells can transition directly into gmgc_head cells (arrows in C and F). This is reflected in the URD tree by the connected segments 25 and 1 (Figures 1.25 L and 1.32 A). (G) Model for linear ZMG/gMGC location dependent changes. Gland cells that are displaced change expression and morphology. Colors of cells correspond to populations depicted in (C,F). Bars show known expression domains for genes plotted in (H). gmgc: granular mucous gland cell, hyp: hypostome, tent: tentacle, zmg: zymogen gland cell. (H) Transdifferentiation resolved as a linear trajectory - URD trajectory plot for selected gland cell genes showing changes in expression along the body axis that recapitulate reported patterns for *HyDkk1/2/4 A/C* (Augustin et al., 2006; Guder et al., 2006), *HyTSR1* (Siebert et al., 2008), *CHIA* (t18356) (Figure 1.32 B-C) and *matrilysin-like* (t32151) (Figure 1.32 D-E). Cells are ordered according to pseudotime with head gland cell states to the left and foot cell states to the right.

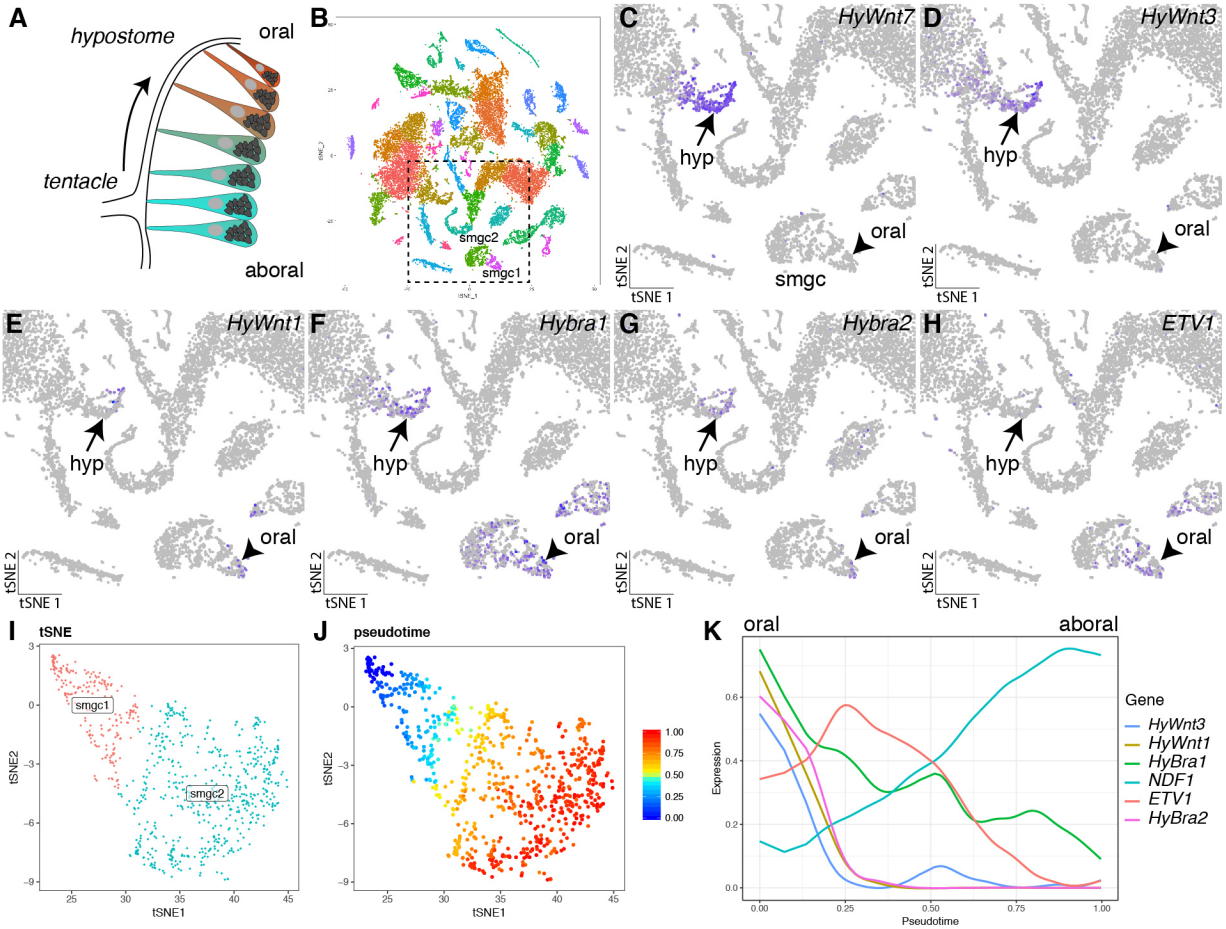


Figure 1.34. URD reconstruction of a differentiation trajectory for spumous mucous gland cells (sMGCs). (A) We hypothesized that there is a progression of states in sMGCs of the *Hydra* head. (B) t-SNE representation for the whole data set. Boxed region is shown in C-H. We explored marker gene expression to identify the oral most cells in the sMGC population. (C-H) Expression plots for selected markers. *HyWnt7*, *HyWnt3*, *HyWnt1*, *Hybra1* and *Hybra2* are expressed in the *Hydra* hypostome ("hyp", arrow) (Bielen et al., 2008; Lengfeld et al., 2009; Technau and Bode, 1999). We find restricted expression in the endodermal epithelial cells of the hypostome for *HyWnt7*, whereas *HyWnt1*, *HyWnt3*, *Hybra1* and *Hybra2* (D-G, arrowhead) are also expressed in sMGCs. Absence of *HyWnt7* expression in sMGCs excludes the possibility of epithelial/gland cell doublets as the source of *HyWnt1*, HyWnt3, *Hybra1* and *Hybra2* expression in the sMGCs. Rather, expression of these genes in sMGCs was likely previously missed due to expression in the surrounding endodermal epithelial cells of the hypostome. (H) Transcription factor *ETV1* (t22116) is expressed in gland cells of the hypostomal tip. (I-K) URD trajectory reconstruction. (I) t-SNE plot for sMGCs. (J) Oral most cells (with *HyWnt1*, *HyWnt3*, *Hybra1* and *Hybra2* expression*)* in cluster smgc1 were selected as root cells (dark blue). Cells were ordered according to their expression and pseudotimes assigned. (K) Trajectory plot for *HyWnt1*, *HyWnt3*, *Hybra1*, *Hybra2*, *ETV1* (t22116) and *NDF1* (t21810) showing expression changes in pseudotime that may correlate to position along the oral aboral axis. Cells are ordered according to pseudotime with putative hypostomal cell states to the left and putative lower head cell states to the right.

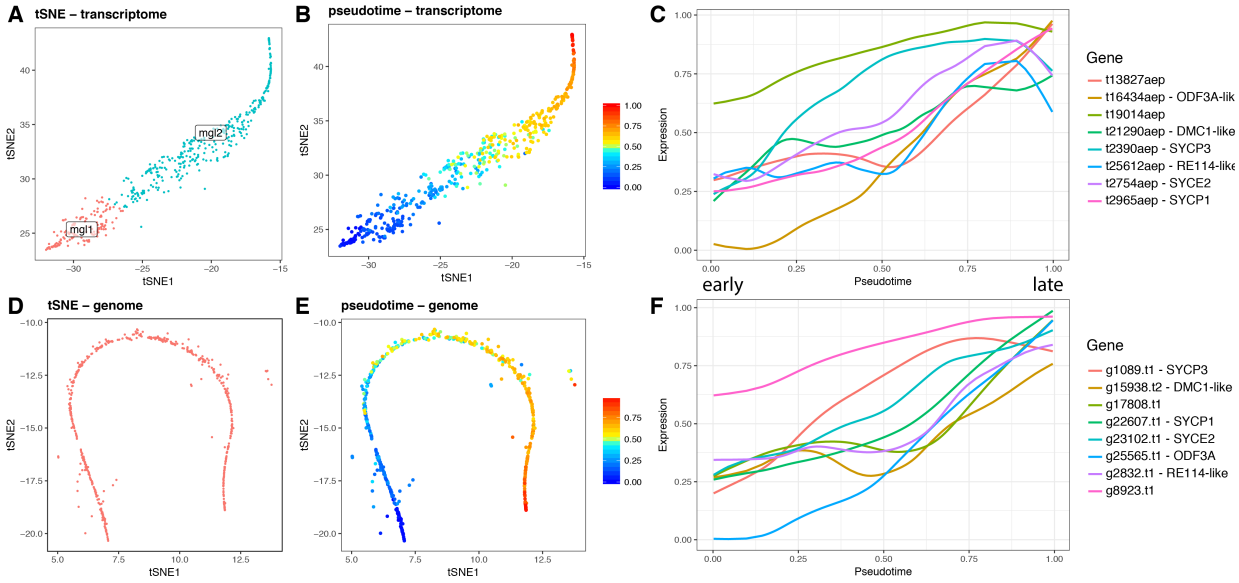


Figure 1.35. URD reconstruction of a trajectory for male germ cells. Trajectories were reconstructed for both transcriptome (A-C) and genome data (D-F) to ensure representation of germline-specific gene expression, since the transcriptome reference was generated from animals not undergoing gametogenesis. Both analyses yielded similar results. (A) t-SNE presentation for male germline clusters (transcriptome). B) Pseudotime reconstruction. (C) Trajectory plots for selected genes that are upregulated during spermatogenesis. (D) t-SNE presentation for the male germline cluster (after mapping reads to the *Hydra* 2.0 genome, see Figure 1.37 and methods). (E) Pseudotime reconstruction for genome data. (F) Trajectory plots for selected genes that are upregulated during spermatogenesis. Genes shown in both trajectories: *SYCP1,3* (Fraune et al., 2012), *SYCE2* (t2754) (this study, Figure 1.36 F), putative sperm tail protein *ODF3A* (t16434aep), meiosis gene *DMC1* (t21290), meiosis gene *RE114* (t25612), uncharacterized gene *t13827/g17808.t1* (this study, Figure 1.36 C), uncharacterized gene *t19014/g8923.t1* (this study, Figure 1.36 D). The expression of synaptonemal complex proteins, *DMC1* (t21290), and *ODF3A* (t16434) peak at the end of the trajectory. The expression of meiotic genes at the end of the trajectory suggests that we captured spermatogonia and spermatocytes, but that spermatids were likely not captured potentially due to their low transcript abundance.

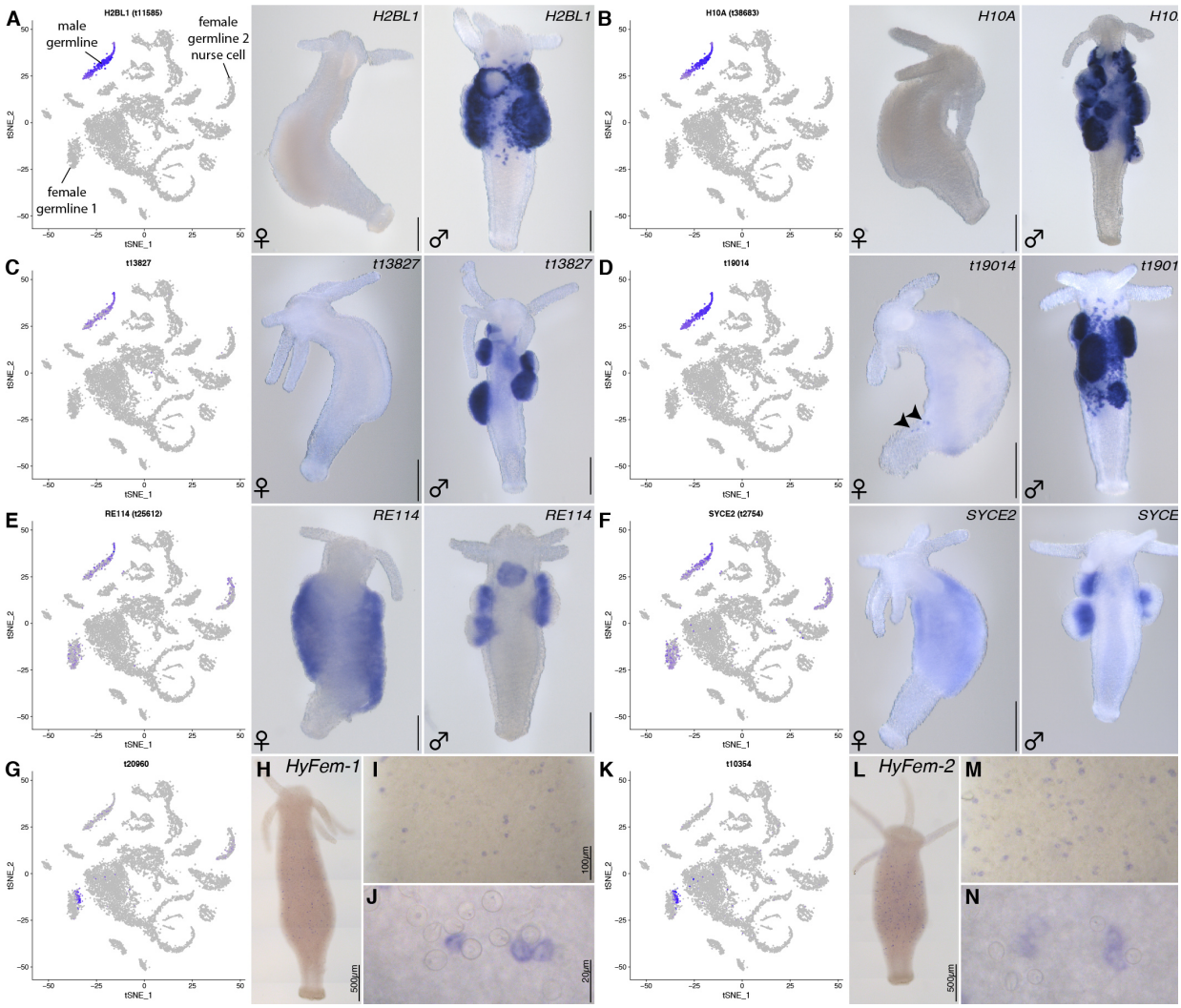


Figure 1.36. Genes expressed in male and female germ cells. (A-F) Expression plots for selected genes (left) with corresponding RNA in situ hybridizations for polyps with developing egg patches (middle), and polyps with testes (right). Labels in A apply to all t-SNE plots in this panel. (A) Male germline specific marker - histone *H2BL1* (t11585aep). (B) Male germline specific marker - histone *H10A* (t38683). (C) Uncharacterized gene *t13827* expressed during spermatogenesis (see Figure 1.35 C). (D) Uncharacterized gene *t19014* expressed during spermatogenesis (Figure 1.35 C). Note positive male cells in specimen with female phenotype (arrows). (E) Meiosis gene *RE114-like* (t25612aep) is expressed during spermatogenesis (Figure 1.35 C) and also expressed in the female germline clusters. (F) *SYCE2* (t2754aep) is expressed during spermatogenesis (Figure 1.35 C) and in cells of the female germline. (G) Expression of newly identified early female germline marker *HvFem-1*. (H-J) *HvFem-1* is expressed in single cells or doublets scattered throughout the body column. (K) Expression of newly identified early female germline marker *HvFem-2*. (L-N) *HvFem-2* in expressed single cells or doublets scattered throughout the body column. *HvFem-2* is expressed in a larger number of cells compared to *HvFem-1*. Scale bars (A-F): 200 µm.

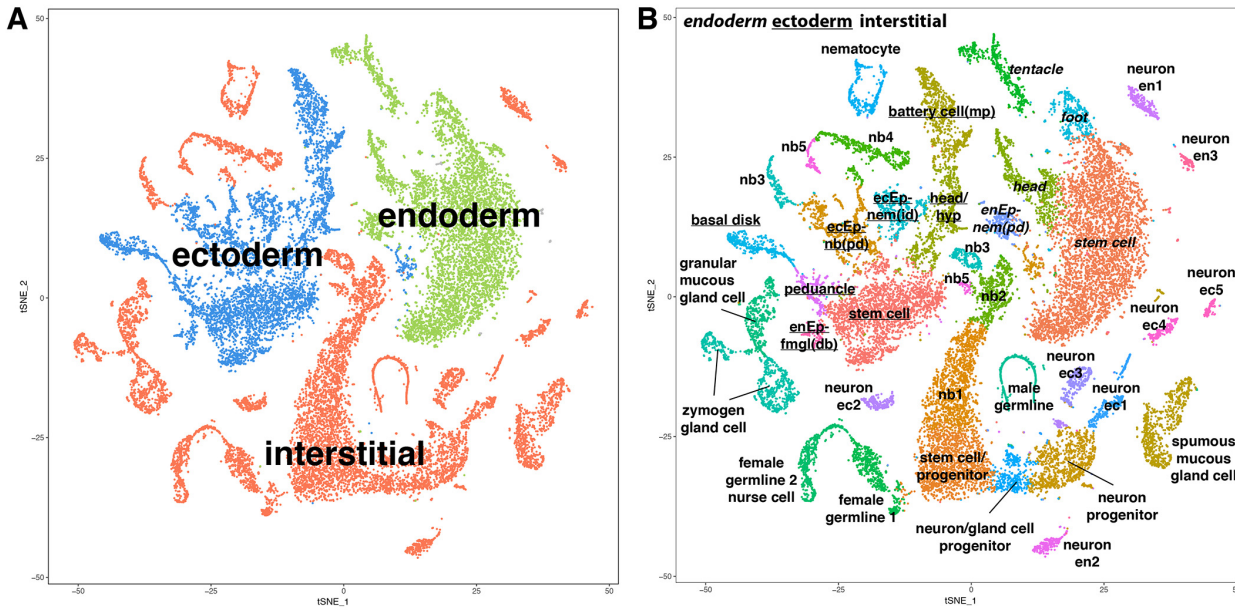


Figure 1.37. Clustering results after mapping Drop-seq reads to the *Hydra* 2.0 genome (*https://research.nhgri.nih.gov/hydra/*). After applying a lower cut-off of 300 genes and 500 UMIs and an upper cut-off of 7,000 genes and 50,000 UMIs, and performing an initial doublet filtering, we obtained 24,458 single cell transcriptomes. (A) t-SNE representation of clustered cells colored by cell lineage. (B) t- SNE representation of clustered cells annotated with cell state. db: suspected doublet, ec: ectodermal, en: endodermal, Ep: epithelial cell, fmgl: female germline, gc: gland cell, hyp: hypostome, id: integration doublet, mp: multiplet, nb: nematoblast, nem: nematocyte, pd: suspected phagocytosis doublet, prog: progenitor. id, mp, pd: are categories of biological doublets.

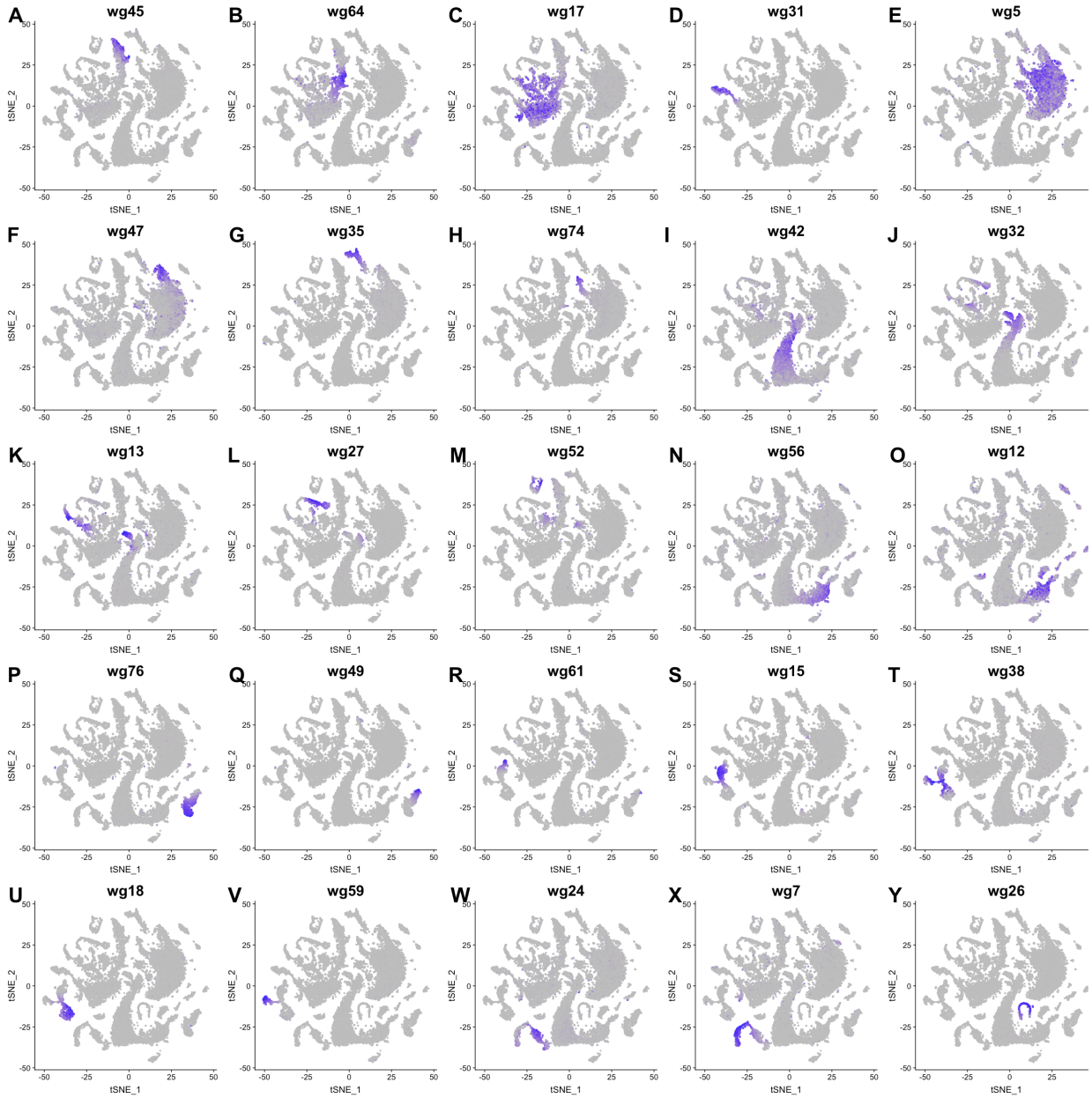


Figure 1.38. Selected metagenes identified in NMF analysis for genome mapped data (wg_K84). A metagene describes a set of genes that are co-expressed in the highlighted cell population. (A) Tentacle ectodermal epithelial cells. This metagene includes transcripts that are expressed in the epithelial cell of a battery cell complex since expression is not found in neuronal or nematocyte cell populations. (B) Ectodermal epithelial cells, head/hypostome. (C) Ectodermal epithelial cells, body column. (D) Ectodermal epithelial cells, basal disk. (E) Endodermal epithelial cells, body column. (F) Endodermal epithelial cells, foot. (G) Endodermal epithelial cells, tentacle. (H) Endodermal epithelial cells, hypostome. (I) Early stage nematoblasts, singletons and phagocytosed. (J) Mid-stage nematoblasts, singletons and phagocytosed. (K,L) Late nematoblasts, singletons and integrated. (M) Mature nematocytes, singletons and integrated. (N) Neuronal cell progenitors. (O) Differentiated neurons and neuronal progenitors. (P,Q) Spumous mucous gland cells. (R) Granular mucous gland cells in the hypostome. (S) Granular mucous gland cells, mid/lower head. (T) Granular mucous gland

cells/zymogen gland cells. (U,V) Zymogen gland cells. (W) Female germline cells. (X) Female germline cells, nurse cells. (Y) Male germline cells. For cluster annotations see Figure 1.37.

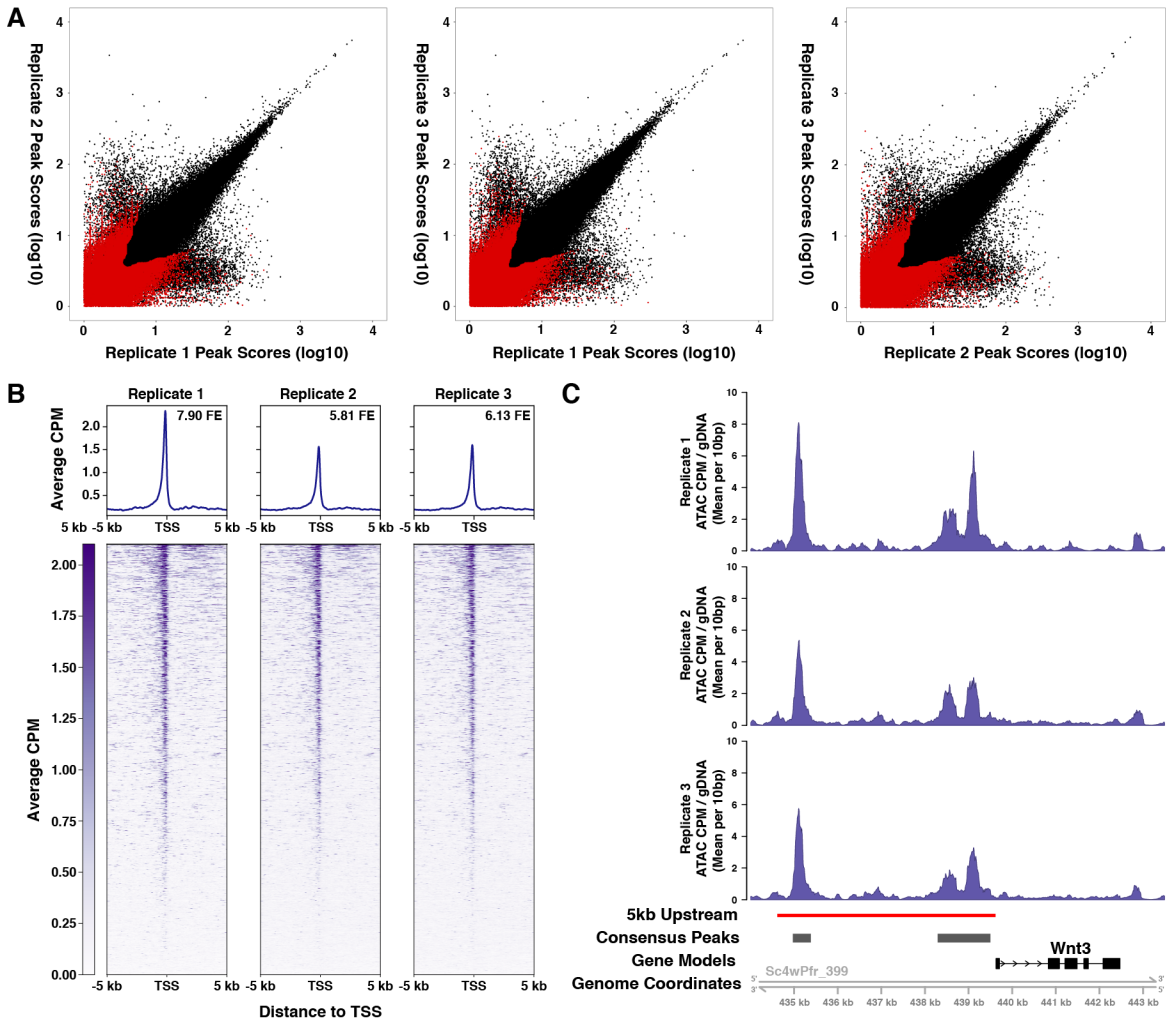


Figure 1.39. ATAC-seq on homeostatic *Hydra*. (A) Pairwise comparisons among the three biological replicates. Replicate log10 peak scores are plotted. Peaks that did not pass the specified IDR (Irreproducible Discovery Rate) threshold of 0.1 are colored red. For downstream analyses, we generated a consensus peak list consisting of all peaks that passed an IDR threshold of 0.1 for at least one pairwise comparison among the three biological replicates. (B) Normalized ATAC signal density in promoters, centered on the TSS. The heatmaps show the normalized ATAC signal in individual sites, sorted by signal intensity. Plotted are the 2000 most highly expressed genes in the single-cell data. FE: fold enrichment, calculated by taking the average read density of the 100bp regions 1kb on either side of the TSS and comparing it to the average read density at the TSS. (C) Example data for gene *HyWnt3a*; visualized is the chromatin accessibility landscape identified in three biological replicates and the consensus peak regions within 5kb upstream of the transcription start site that were considered in enrichment analyses. Replicate tracks for chromatin accessibility and consensus peaks can be visualized at the *Hydra* 2.0 Web Portal (*https://research.nhgri.nih.gov/hydra/*).

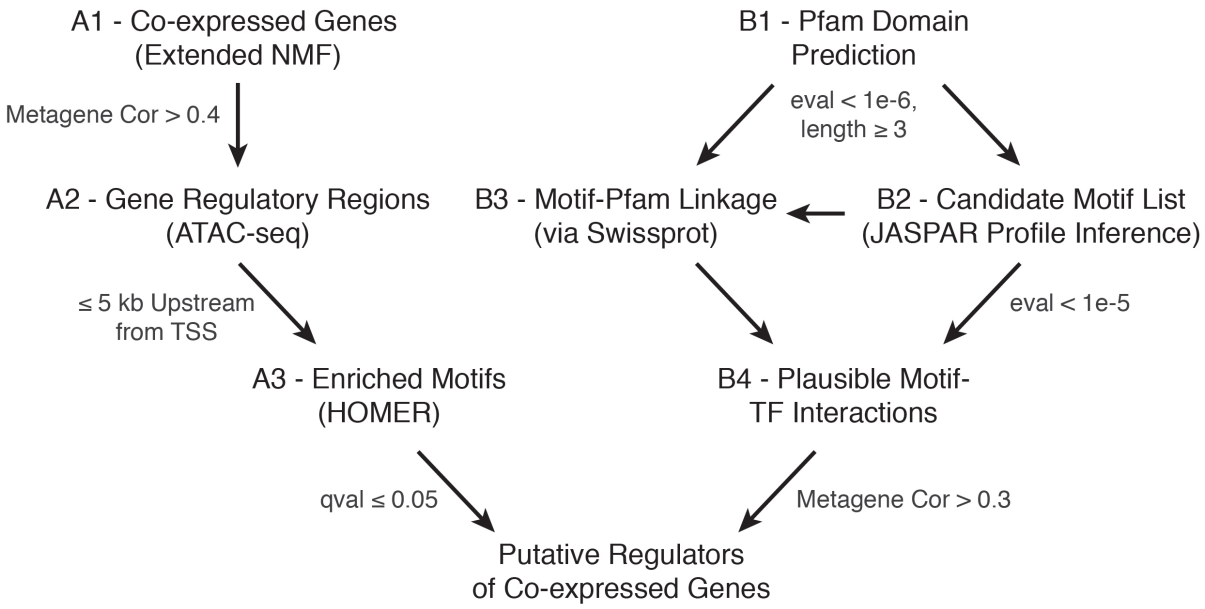


Figure 1.40. Workflow for the identification of candidate transcriptional regulators. (A1) Co-expressed genes (metagenes) were identified using NMF. Genes were considered members of a metagene if their expression had a correlation score > 0.4 when compared against all non-zero cell scores for that metagene. (A2) Regulatory regions associated with metagenes were identified using ATAC-seq data. All biologically reproducible peaks that were within 5kb upstream of metagene-associated genes were used for downstream motif enrichment analysis. (A3) HOMER was used to test for motif enrichment in metagene- associated regulatory sequences with a cutoff of q-value (adjusted p-value) ≤ 0.05. (B1) Conserved domains were identified using HMMR and the Pfam database, with only alignments longer than 3 amino acids and an E-value < 1e-6 being used for downstream analysis. Putative transcription factors (TFs) were identified based on the presence of a Pfam domain predicted to have DNA binding activity. (B2) Putative TFs were linked to binding motifs using the Profile Inference tool provided by the JASPAR database with an E-value cutoff of < 1e-5. (B3) The binding motifs provided by Profile Inference were linked to additional putative TFs based on those TFs having a Pfam annotation for the domain that binds the motif in question. (B4) Results from B2 and B3 were compiled into a table linking motifs to their potential binding partners in the *Hydra vulgaris* (strain 105) genome. A putative regulator of gene co-expression needed to fulfill two criteria: 1) its expression needed to correlate with the expression domain of a metagene (correlation score > 0.3) and 2) a motif that could be bound by the regulator needed to be enriched in the upstream regulatory regions of the genes within that metagene. Metagenes and candidate regulators are listed in Table 1.5.

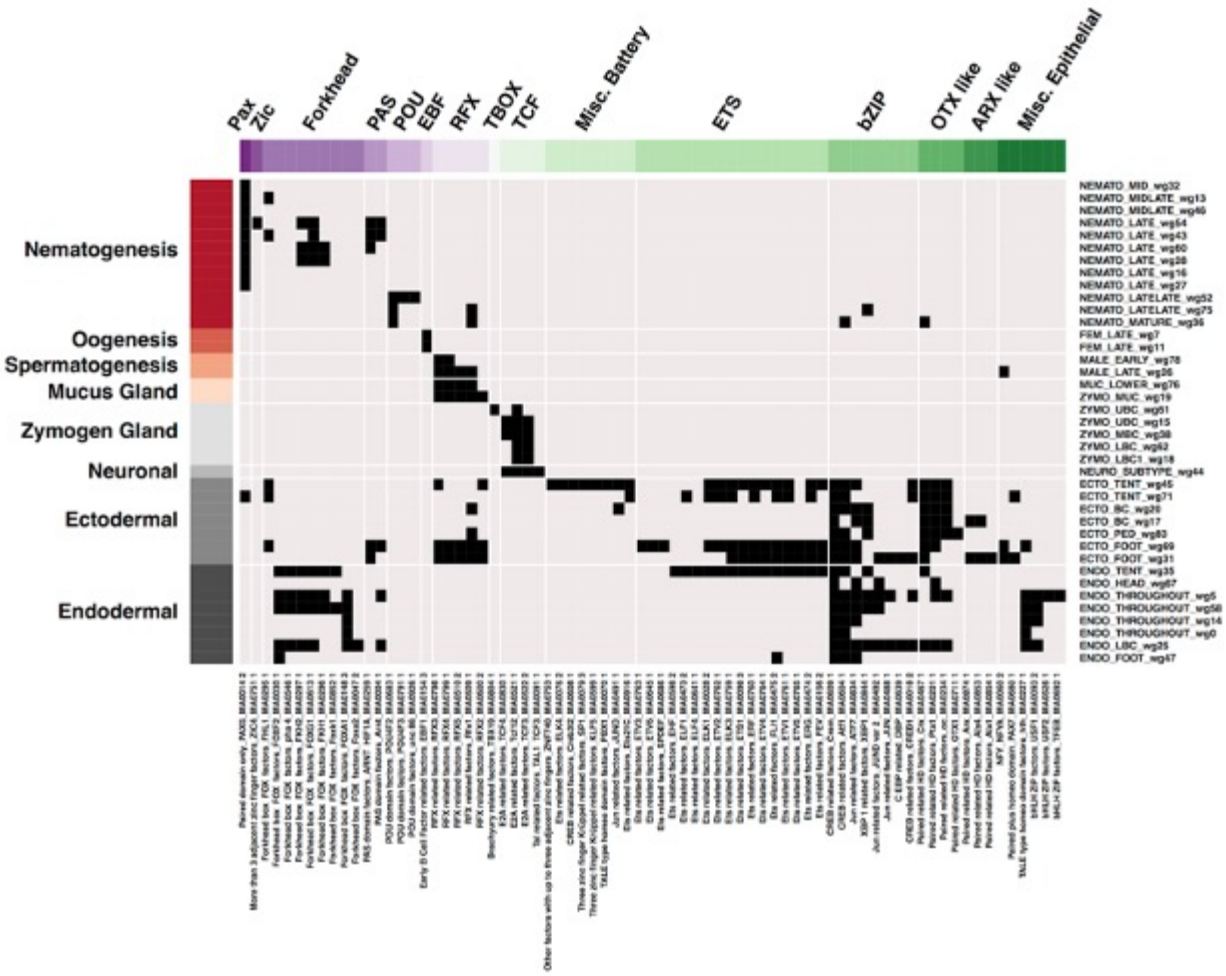


Figure 1.41. Motif enrichment analysis for gene modules. Motifs found enriched in open chromatin of 5' cis- regulatory regions of co-expressed gene sets (metagenes) in particular cell states. Rows represent metagenes (right) and these are organized by cell state (left). Columns represent individual JASPAR transcription factor (TF) binding motifs. Column names include the TF family name, the short name, and the JASPAR motif ID pulled from the JASPAR database. Metagenes were annotated with cell states, location or differentiation phase. bc: body column, ecto: ectodermal epithelial cells, endo: endodermal epithelial cells, fmgl2: female germline nurse cells, gmgc: granular mucous gland cell, mgc: mucous gland cell, mgl: male germline, nb: nematoblast, nem: differentiated nematocytes, ped: peduncle, smgc: spumous gland cell, ubc: upper body column, zmg: zymogen gland cell. Metagenes can be visualized at the Broad Single-Cell Portal. See Materials and Methods for detail.

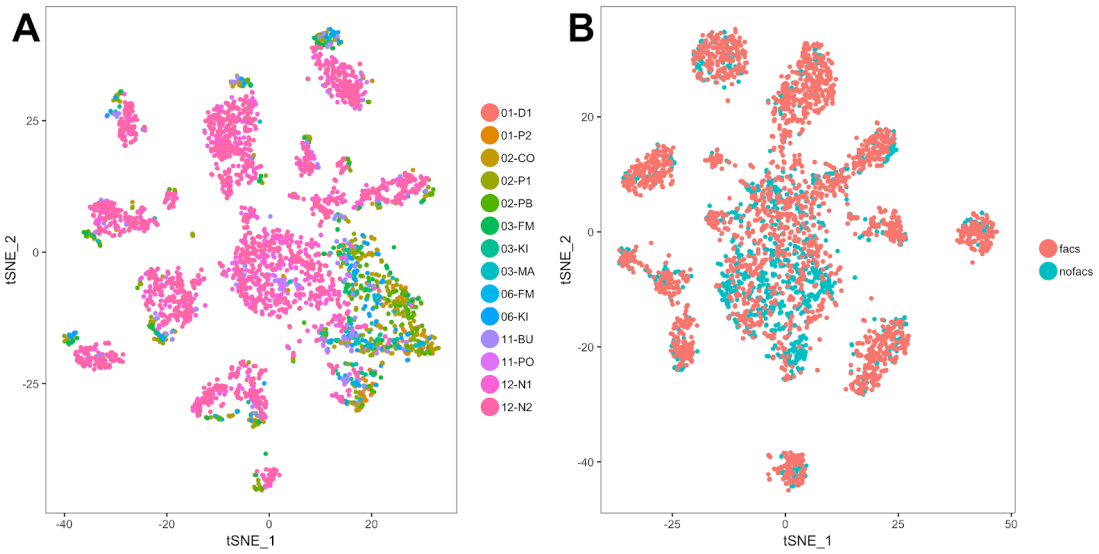


Figure 1.42. Subclustering of neuronal cells. (A) t-SNE representation after clustering without batch correction revealed library specific (batch) effects introduced by adding FACS to the workflow. Cells from libraries 12-N1 and 12-N2 were enriched for neuronal cells using FACS (Table 1.1). (B) Cells could be integrated following the approach by Butler et al. (2018). See supplementary analysis SA05 for details. t-SNE representation after integration of cells from sorted (FACS) and unsorted libraries.

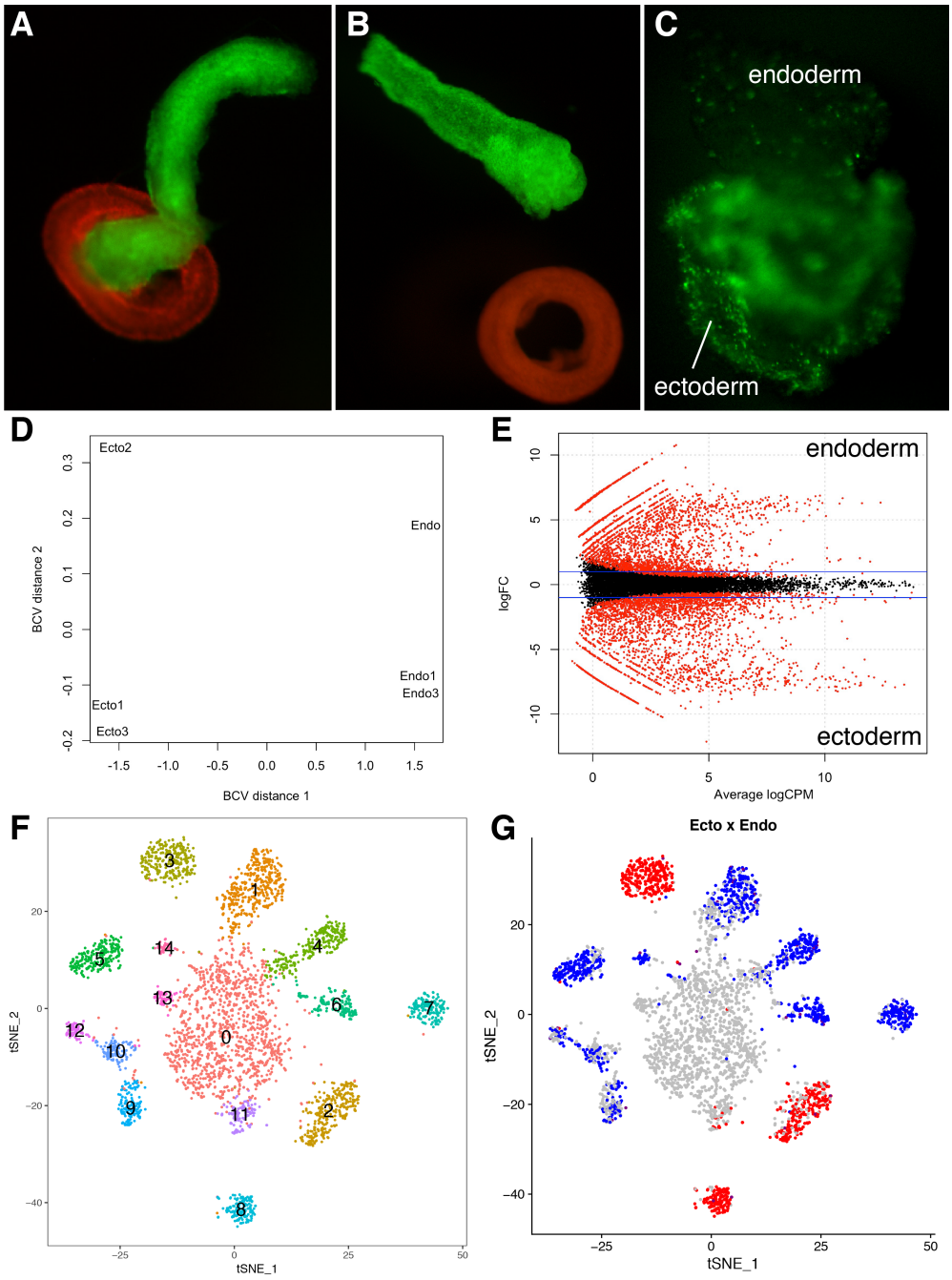


Figure 1.43. Epithelium-specific differential gene expression and placement of neuronal cells. (A,B) Endoderm and ectoderm of the body column were separated after removing head and foot tissue for subsequent epithelium specific RNA-seq. The Inverse Watermelon transgenic line was used, which expresses RFP in ectodermal epithelial cells and GFP in endodermal epithelial cells, which allowed us to assess the quality of our tissue separation. (C) Tissue separation using the nGreen transgenic line, in which GFP is predominantly expressed in neurons, demonstrates that neurons stay associated with the epithelial tissue after separation. (D) Demonstrating the paired nature of three endoderm and three ectoderm 3'Tag- Seq libraries. Distances between samples correspond to the leading biological coefficient of variation (BCV). (E) Differential gene expression analysis results. Genes that are significantly differentially

expressed are colored in red. Positive fold-change indicates enrichment in the endodermal epithelium. x-axis shows log-scaled abundance (counts per million). (F) Neuronal subclustering. Neuronal cells were scored for genes differentially expressed between epithelia. (G) Cells with high scores for ectodermal genes colored in blue, cells with high scores for endodermal genes colored in red. Cells from clusters 2, 3, and 8 are neurons located in the endodermal epithelium. Cells from clusters 1, 4, 5, 6, 7, 9, 10, 12 are neurons located in the ectodermal epithelium. Cluster 11 is positive for endodermal genes (red cells) suggesting endodermal progenitors since they also express progenitor markers like *Myb*/*Myc3*. Additional expression of differentiation markers such as *COMA* (t2163, Figure 1.27 E,J,O) suggests gland cell or neuron fate.

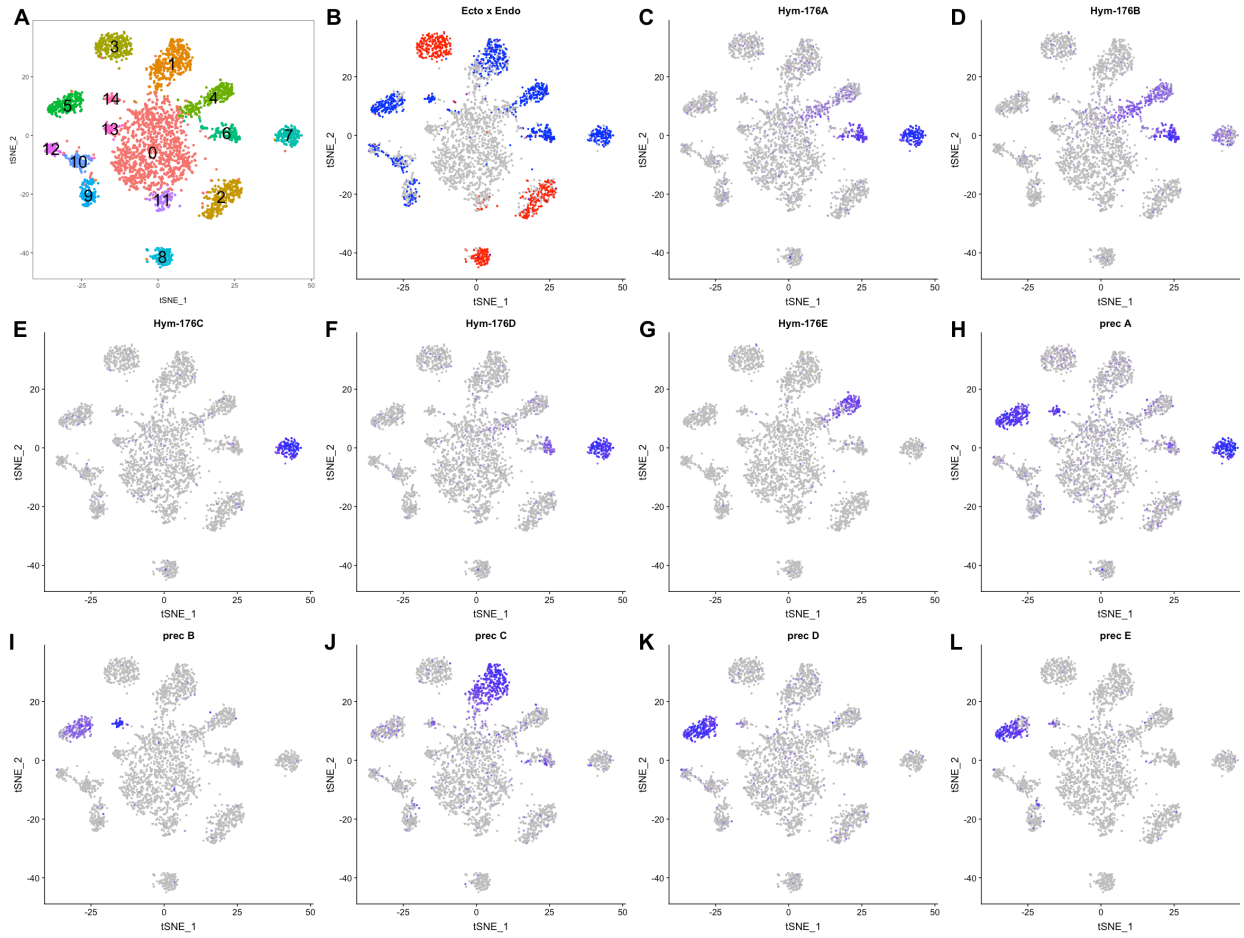


Figure 1.44. Neuronal markers used in cluster annotation. (A) Neuronal subclustering. (B) Neuronal cells colored by epithelial origin. Red - endoderm, blue - ectoderm. C-K) Published expression of *Hym-176* paralogs and Rfamide precursors were used to annotate neuronal clusters. Format of legend: transcript name; published expression pattern (cluster number) (C) *Hym-176A;* ectodermal ganglion neurons in the head/tentacles (cl 4), body column (cl 6) and peduncle (cl 7) (Noro et al., 2019; Yum et al., 1998). (D) *Hym-176B*; ectodermal ganglion neurons in the head/tentacles (cl 4) and body column (cl 6) (Noro et al., 2019), (E,F) *Hym-176C* and *Hym-176D*; ectodermal ganglion neurons in the peduncle (cl 7) (Noro et al., 2019), (G) Hym-176E; ectodermal ganglion neurons in the head/tentacles (cl 4) (Noro et al., 2019). (H) *RFamide preprohormone A*; ectodermal sensory and ganglion neurons in the tentacle (cl 5), head/hypostome (cl 14), and peduncle (cl 7) (Darmer et al., 1998). (I) *RFamide preprohormone B;* ectodermal sensory neurons in the tentacles (cl 5) and the head/hypostome (cl 14) (Darmer et al., 1998). (J) *RFamide preprohormone C;* ectodermal sensory neurons in the tentacles (Darmer et al., 1998). (K) *RFamide preprohormone D;* ectodermal sensory neurons in the tentacles (cl 5) (Hayakawa et al., 2004). (L) We identified a novel *RFamide preprohormone E* expressed in cells of cluster 5. In situ hybridization revealed expression in ectodermal sensory neurons in the tentacles (Figure 1.46 E-G)**.**

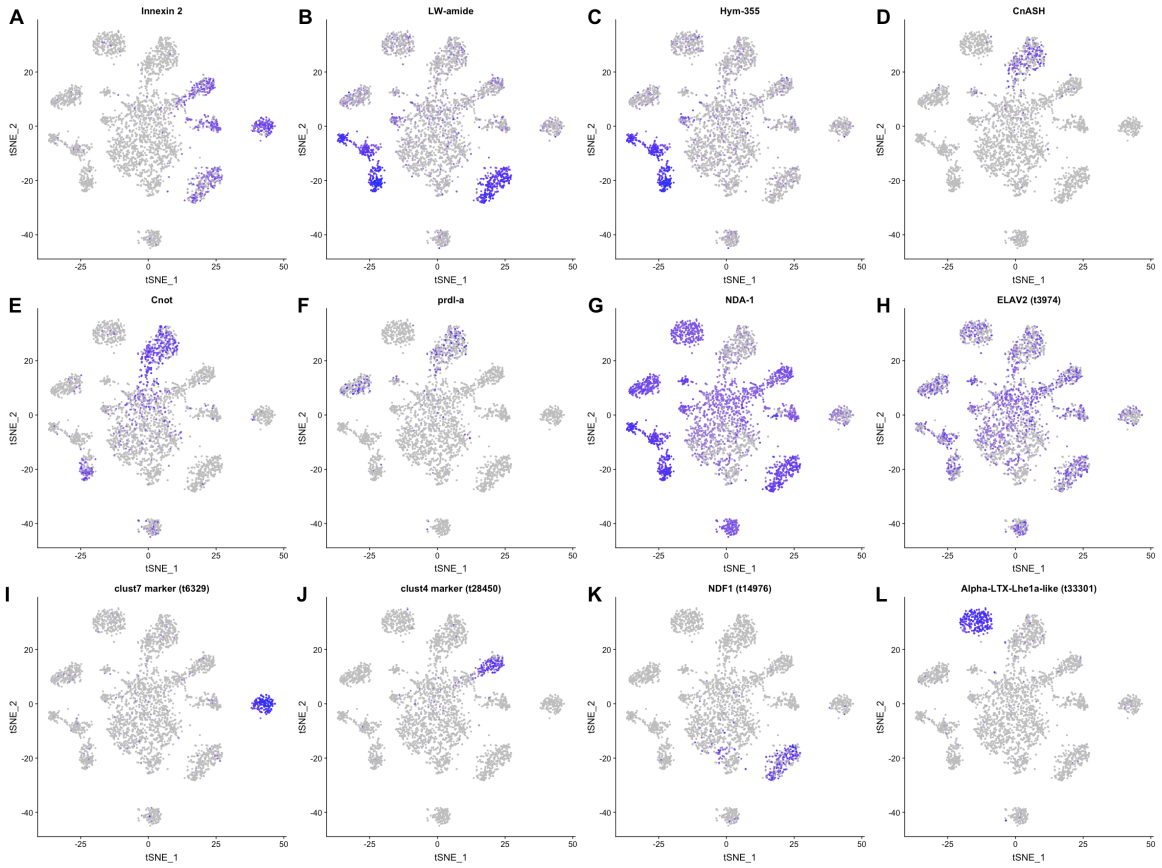


Figure 1.45. Neuronal markers used in cluster annotation. Neuronal genes with previously published expression patterns (A-G) and neuronal genes identified in this study (H-L) were used to annotate neuronal clusters. Format of legend in A-F: transcript name; published expression pattern (cluster number). For clustering numbering see Figure 1.44 A. (A) *Innexin 2;* ectodermal ganglion cells of the peduncle (cl7) (Takaku et al., 2014). The data suggests a wider expression in ectodermal neurons of the body column (cl 6) and head (cl 4) and a subset of endodermal neurons (cl 2). (B) *LW-amide;* expression in ectodermal ganglion neurons in the tentacles/head (cl 9), body column (cl 10), and basal disk (cl 12) (Mitgutsch et al., 1999). In addition, RNA in situ hybridization confirmed expression in endodermal neurons (cl 2) (Figure 1.46 H). (C) *Hym-355;* ectodermal ganglion neurons in the tentacles/head (cl 9), body column (cl 10), and basal disk (cl 12) (Takahashi et al., 2000). Cells of cl 13 are closely associated with the progenitors in cl 0, show expression of progenitor marker *Myc3* but also elevated expression of *LW-amide*/*Hym-355* and are tentatively annotated as "ec3_progenitor" (see Figure 1.6 A). (D) *CnASH;* ectodermal sensory neurons in the tentacles (cl 1) (Hayakawa et al., 2004). (E) *Cnot;* ectodermal sensory neurons in the tentacles (cl 1) (Galliot et al., 2009). The data suggest a broader expression in a subset of ectodermal ganglion neurons (9) and a subset of endodermal neurons (cl 8). (F) *prdl-a;* expression in ectodermal sensory neurons in the tentacles (cl 1, 5) (Miljkovic-Licina et al., 2004). (G) *NDA-1* is broadly expressed (with the exception of cluster 1) and has reported expression in both sensory and ganglion neurons throughout the entire body (Augustin et al., 2017). (H) A *Hydra* Elav homolog, ELAV2 (t3974), is expressed in all neuronal clusters and in neuronal progenitors. In situ hybridization revealed strongest expression in developing buds and tentacle bases (Figure 1.46 L-M). (I) *t6329* is expressed in cluster 7 and in situ hybridization confirms expression in ectodermal peduncle neurons (cl 7) (Figure 1.46 C-D).

(J) *t28450* is expressed in cluster 4 and in situ hybridization confirms expression in ectodermal tentacle neurons (Figure 1.46 A-B). (K) *NDF1* (t14976) is expressed in cluster 2 and in situ hybridization and a GFP reporter line confirms expression in endodermal ganglion neurons throughout the body with the exception of the tentacles (Figure 1.6 B-C, Figure 1.46 I-K,N,O). (L) *Alpha-LTX-Lhe1a-like* (t33301) is expressed in cluster 3 and a GFP reporter confirms expression in endodermal sensory neurons throughout the body column (Figure 1.6 B, D-E).

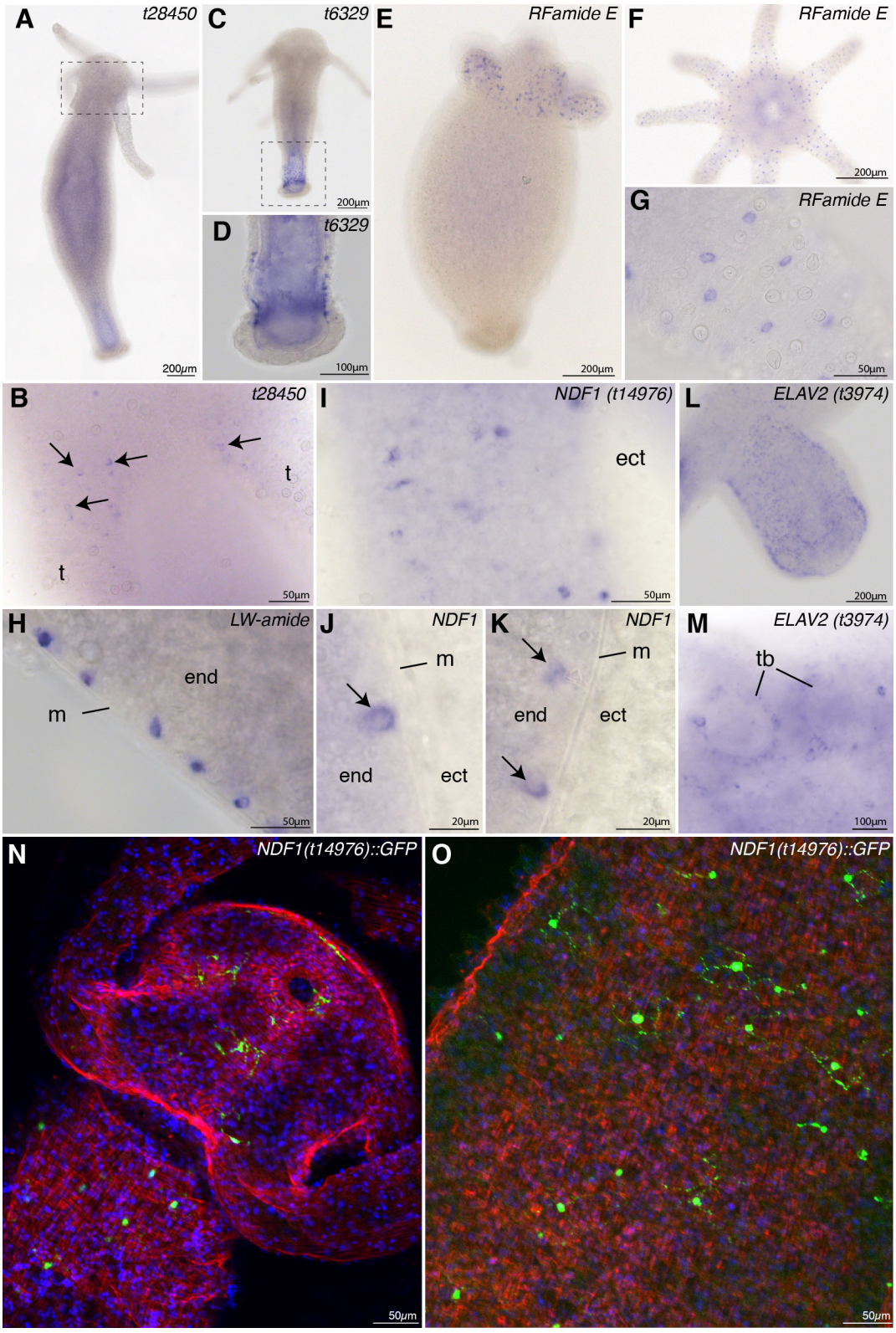


Figure 1.46. Validation of new neuronal markers using RNA in situ hybridization and transgenesis. (A,B) *t28450* is expressed in tentacle neurons found in cluster 4 (Figure 1.45 J) . (B) Close-up of box in A showing expression in neurons of the tentacle base (cl 4). (C,D) *t6329*

is expressed in peduncle neurons found in cluster 7 (Figure 1.45). (D) Close-up of peduncle shown in C demonstrating ectodermal localization of the signal. (E-G) *RFamide preprohormone E* with expression in tentacle neurons (cluster 5) (Figure 1.44 L). (H) LW- amide (*79*) expression in endodermal neurons (cluster 2), ectodermal tissue was removed (Figure 1.45 B). (I-K) *NDF1* (t14976) is expressed in endodermal neurons of the body column (cluster 2) (Figure 1.45 K). (J, K) *NDF1* close-ups, demonstrating endodermal localization of the signal. (L-M) *ELAV2* (t3974) expression is detected most clearly in developing buds (L) and cells at the tentacle base (M) consistent with neuronal progenitor expression (Figure 1.45 H), (N-O) NDF1 (t14976)::GFP mosaic transgenic line with GFP positive ganglion cells in the head (N) and within the body column (O) (cl3, Figure 1.45 K). ect: ectoderm, end: endoderm, tb: tentacle base.

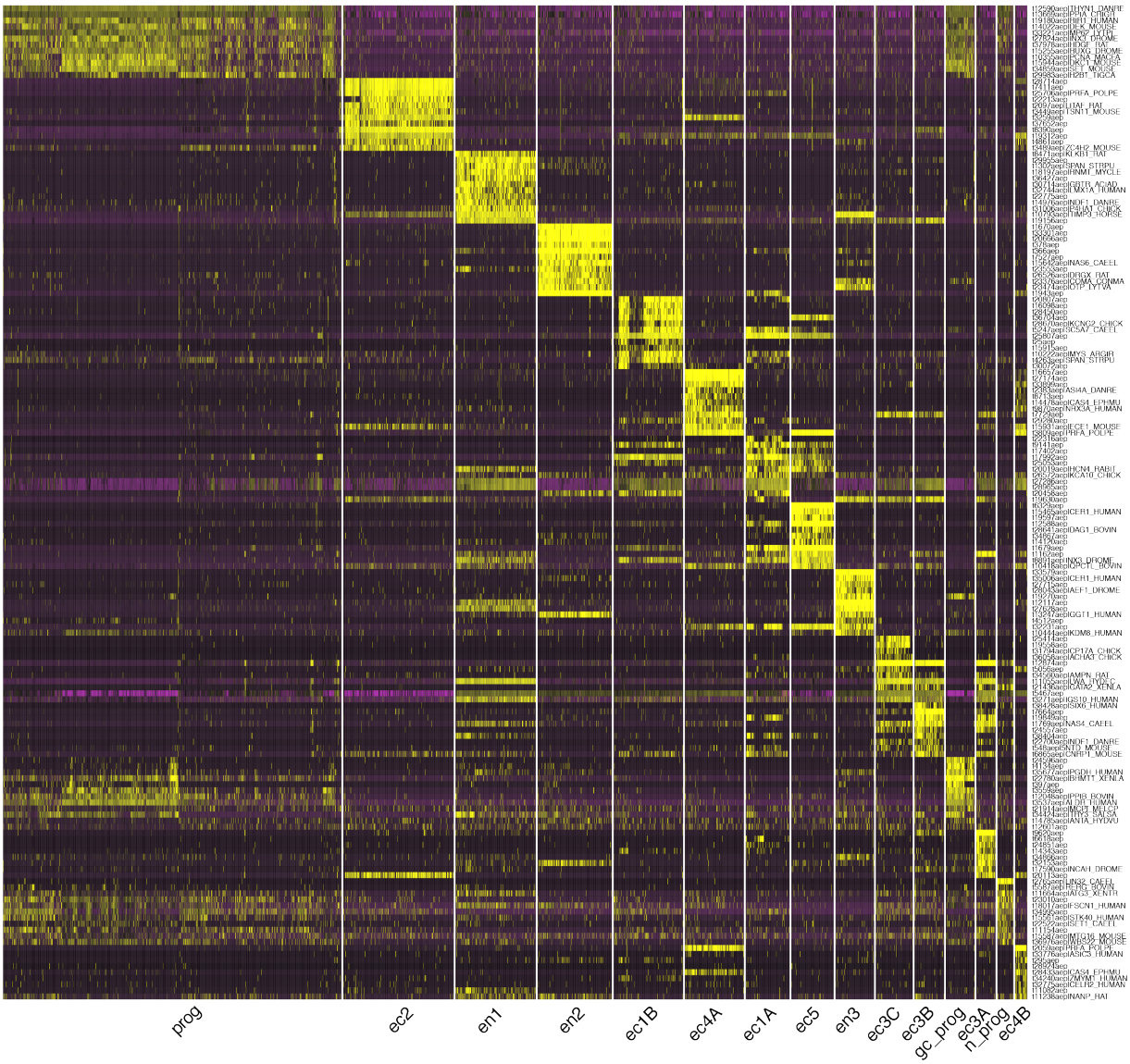


Figure 1.47. Heatmap showing the top twelve markers for each neuronal cluster. The analysis highlighted gene markers (rows) that are expressed in specific clusters (columns), e.g. *LWamide* (t11055) is expressed in clusters ec3A, ec3B, ec3C, and en1. Specific markers exist for each of these *LWamide*-positive clusters. Neuron specific expression should be tested by visualizing expression using the full data set

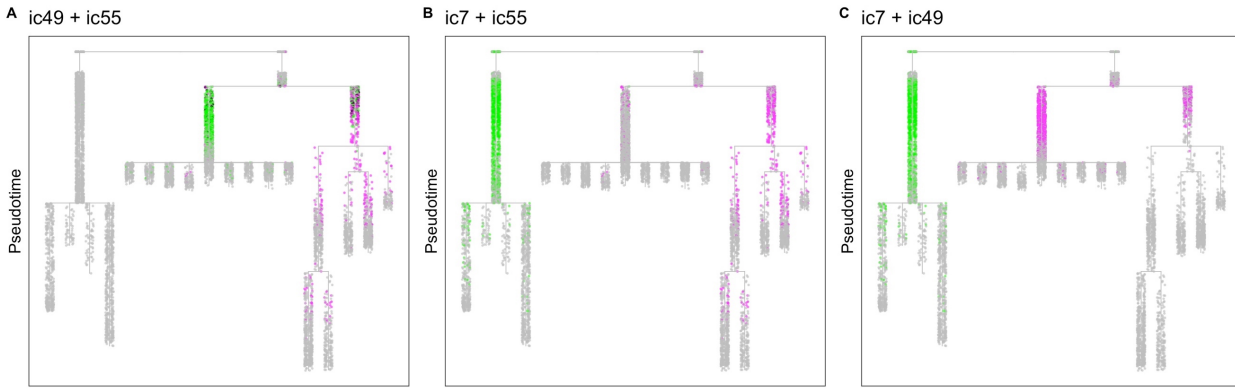
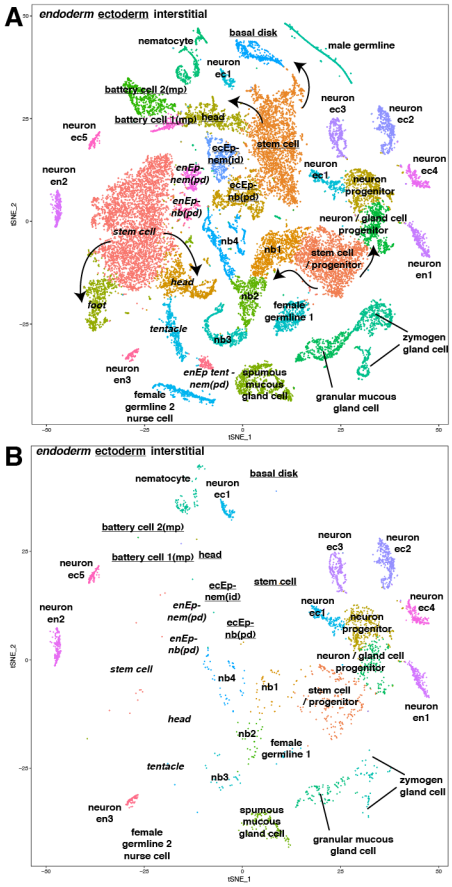


Figure 1.48. Metagene co-expression analysis supports similarities between neurogenesis and gland cell differentiation. NMF, as an unbiased approach, identifies local patterns in the data based on the expression of many genes. For the interstitial data set we identified gene modules specific to the processes of nematogenesis (metagene ic7), neurogenesis (metagene ic49), and gland cell differentiation (metagene ic55) (see Figure 1.25). Co-expression analysis identifies cells that express both the neurogenesis and the gland cell differentiation metagene suggesting that both differentiation programs are simultaneously activated in a single cell. Similar levels of co-expression are not observed when visualizing nematogenesis metagene expression together with neurogenesis or gland cell differentiation metagene expression. Co-expressing cells are visualized in black. A minimum module score cutoff of 0.20 was used in all comparisons. (A) Co-expression of ic49 (green, neurogenesis) and ic55 (magenta, gland cell differentiation). 36 shared genes were found between metagenes considering genes with scores > 0.5. (B) Co-expression of ic7 (green, nematogenesis) and ic55 (magenta, gland cell differentiation). 7 shared genes were found between metagenes (gene scores > 0.5). (C) Co-expression of ic7 (green, nematogenesis) and ic49 (magenta, neurogenesis). 14 shared genes were found between metagenes (gene scores > 0.5). NMF results and information on genes that are included in the metagenes are provided in the accompanying git repository.



C

	cells	medianGene	medianUMI
i_neuron_progenitor	384	793	1498
i_neuron_ec2	381	457	862
i_neuron_ec1	367	490	787
i_neuron_ec3	266	437	869
i_neuron_en1	238	506	871
i_neuron_ec4	215	456	839
i_stem_cell/progenitor	206	1154	2886
i_neuron_en2	196	488	918
i_neuron/gland_cell_progenitor	171	718	1432
i_spumous_mucous_gland_cell	134	794	2096
i_neuron_ec5	118	510	1198
i_granular_mucous_gland_cell	102	846	2218
i_nematocyte	99	665	1121
i_neuron_en3	99	541	1079
i_zymogen_gland_cell	56	1411	6797
i_nb4	51	416	975
i_nb1	45	905	2214
i_nb2	32	836	1768
i_nb3	22	828	2604
i_female_germline1	12	2489	7018
enEp_stem_cell	11	2657	7599
enEp_tentacle	3	3553	15369
ecEp_nb(pd)	2	4032	24592
ecEp_stem_cell	1	1373	3466
ecEp_head	1	2279	6486
enEp_foot	1	740	1424
ecEp_battery_cell2(mp)	1	4964	22784
ecEp_basal_disk	1	431	728
enEp_nem(pd)	1	1940	5289
enEp_nb(pd)	1	2391	6963

Figure 1.49. Cells recovered in neuronal FACS libraries. The transgenic line (nGreen) expresses the GFP transgene predominantly in cells of the neuronal lineage. 75.7% of the 3,218 cells that are retained after filtering are neuronal progenitors or differentiated neurons. (A) t-SNE representation for the whole data set. (B) t-SNE for subset of cells collected following FACS (libraries 12-N1, 12-N2). (C) Number of cells recovered for each state, median gene and median UMI numbers. i: cells of the interstitial lineage, ecEp: ectodermal epithelial cell, enEP: endodermal epithelial cell, mp: multiplet, nb: nematoblast, pd: suspected phagocytosis doublet.

Library	Strain 1	# Animals Strain 1	Strain 2	# Animals Strain 2	Dissociation Medium	Cell Suspension Mediuim	Lysis Buffer (H20)
01-D1	AEP	50	AEP (nGreen)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
01-P2	AEP	50	AEP (nGreen)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
02-CO	AEP	45	AEP (nGreen)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
02-P1	AEP	45	AEP (nGreen)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
02-PB	AEP	45	AEP (nGreen)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
03-FM	AEP (nGreen)	40	AEP (nGreen) - egg patch	10	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
03-KI	AEP	45	AEP (nanos::gfp)	5	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
03-MA	AEP (nGreen)	40	AEP (nGreen) - testes	10	HCM	HyPBS-0.05% BSA	270mM NaCl, 5.2mM KCl
06-FM	AEP - egg patch	40	-	-	HDM	HyPBS(iso)-0.05% BSA	254mM NaCl
06-KI	AEP	40	-	-	HDM	HDM, 0.01% BSA	254mM NaCl
06-MA	AEP - testes	40	-	-	HDM	HyPBS(iso)-0.05% BSA	254mM NaCl
11-BU	AEP - buds	30	-	-	HDM	HDM, 0.01% BSA	254mM NaCl
11-PO	AEP	40	-	-	HDM	HDM, 0.01% BSA	254mM NaCl
12-N1	AEP (nGreen)	40	-	-	HDM	HDM, 0.01% BSA	254mM NaCl
12-N2	AEP (nGreen)	40	-	-	HDM	HDM, 0.01% BSA	254mM NaCl

Table 1.1: *Hydra* strains and media used in dissociations. Fifteen Drop-seq libraries were generated. Libraries 01-11 were generated using whole animal cell suspensions. Libraries 01-06 and library 11-PO were generated using polyps without apparent buds. Libraries 12-N1 and 12-N1 were generated using cell suspensions that were enriched for neuronal cells using fluorescence activated cell sorting (FACS). The table shows the number of animals and strains used in the experiments, the medium used in the dissociations, the medium the cells were resuspended in prior to performing Drop-Seq and the salt solution that replaced water in the Drop-seq lysis buffer. Salt was added to keep salt concentration in the droplets close to the conditions described in Macosko et al. (Evan Z. Macosko et al., 2015). AEP: *Hydra vulgaris* strain AEP, courtesy of Thomas C.G. Bosch (Kiel). AEP (nanos::gfp): *Hydra vulgaris* AEP strain expressing GFP driven by the Cnnos1 promoter (Hemmrich et al., 2012). AEP (nGreen): *Hydra vulgaris* AEP strain expressing GFP driven by actin promoter courtesy of Rob Steele (see Figure 1.49**)**. buds: polyps were bearing buds of various stages, egg patch: polyps were undergoing oogenesis, testes: polyps were bearing testes, HCM: *Hydra* culture medium, HDM: *Hydra* dissociation medium.

Library	cells/µl	beads/µl	Droplet Size (nl)	Cell Occupancy	Multiplet Rate (%)	STAMPs in PCR	# Cycles in PCR	Recovery (%)
01-D1	100	130	0.98	0.048	2.42	100	12	40.2
01-P2	120	130	0.98	0.057	2.90	100	13	41.9
02-CO	180	130	0.98	0.084	4.32	296	15	59.9
02-P1	180	130	0.98	0.084	4.32	296	12	70
02-PB	180	130	0.98	0.084	4.32	296	12	53.3
03-FM	180	130	0.98	0.084	4.32	184	14	29.2
03-KI	180	130	0.98	0.084	4.32	148	13	60.5
03-MA	180	130	0.98	0.084	4.32	148	14	37.8
06-FM	250	185	0.70	0.084	4.31	145	12	37.1
06-KI	205	166	0.78	0.077	3.94	133	12	66.6
06-MA	250	185	0.70	0.084	4.31	145	12	49.8
11-BU	160	167	0.78	0.060	3.09	130	14	86.8
11-PO	160	167	0.78	0.060	3.09	130	16	51.9
12-N1	113	167	0.78	0.043	2.19	200	16	51.9
12-N2	135	185	0.70	0.046	2.34	200	15	58.4

Table 1.2: Drop-seq. The table holds cell and bead concentration, droplet sizes generated by the microfluidics device, cell occupancy, Drop-seq multiplet rate (expected technical multiplet rate based on cell concentration loaded and assuming a pure single cell suspension), STAMPS processed per 50µl PCR reaction, number of cycles used in the library amplification. The percent recovery indicates the ratio of number of single cell transcriptomes that were obtained (using the inflection point after plotting the cumulative fraction of reads per cell against the cell barcodes) versus number of anticipated STAMPS based on cell occupancy. Since recovery rates were low in pilot libraries we hypothesized that not all counted cells were captured as cells in Drop-seq and cell concentrations were increased (see considerations in “Bead and cell concentration”).

Library	Cells	Median Genes	Median UMI
01-D1	1023	3400	15235
01-P2	1256	3090	10534
02-CO	2329	1941	5174
02-P1	3343	2413	7205
02-PB	1562	2596	8006
03-FM	886	2644	10309
03-KI	1958	2460	9914
03-MA	702	3056	12474
06-FM	1122	3722	13932
06-KI	1937	2075	7318
06-MA	384	1621	3466
11-BU	3207	1342	3281
11-PO	2058	955	2350
12-N1	1264	570	1085
12-N2	1953	557	1079

Table 1.3. Number of single cell transcriptomes (cells), median genes and median UMIs per cell considering cells from all libraries. Metrics for cells after applying gene and UMI cut-offs (>300 <7k genes and >500UMI <50k UMIs) and initial doublet filtering (24,985 cells, transcriptome reference).

Library	Cells	Median Genes	Median UMI
i_neuron_ec2	442	492	958
i_neuron_ec4	267	501	920
i_neuron_ec3	366	512	1042
i_neuron_ec1	478	545	874
i_neuron_ec5	160	552	1316
i_neuron_en1	311	556	1012
i_neuron_en2	287	574	1128
i_neuron_en3	143	628	1247
i_nb4	701	695	1636
i_nematocyte	549	925	1767
i_nb3	529	1112	3607
i_granular_mucous_gland_cell	560	1132	3492
i_neuron_progenitor	717	1203	2537
i_nb2	608	1229	2770
i_spumous_mucous_gland_cell	869	1238	3601
i_zymogen_gland_cell	783	1326	8250
i_neuron/gland_cell_progenitor	567	1780	4608
enEp_tentacle	458	1795	4921
i_male_germline	535	1800	4072
i_nb1	902	1938	5404
ecEp_basal_disk	452	1942	5957
enEp_tent(pd)	134	2122	6194
i_stem_cell/progenitor	1879	2153	6738
ecEp_battery_cell1(mp)	200	2330	6357
enEp_head	825	2374	6783
enEp_foot	659	2571	7971
i_female_germline1	521	2653	7014
enEp_stem_cell	4005	2760	9361
ecEp_stem_cell	2708	2947	10518
enEp_nem(pd)	257	2968	9732
ecEp_head	695	3001	10351
enEp_nb(pd)	143	3065	10704
ecEp_battery_cell2(mp)	608	3153	9921
ecEp_nem(id)	449	3492	12719
ecEp_nb(pd)	759	3660	14648
i_female_germline2_nurse	458	4640	16686

Table 1.4. Median genes/UMIs per cell per state (all libraries, transcriptome reference). Metrics for cell states after applying gene and UMI cut-offs (>300 <7k genes and >500UMI <50k UMIs, sorted by number of genes detected) and initial doublet filtering (24,985 cells). For clustering see Figure 1.1 F. Neurons presented the lowest detected number of genes and UMIs, while female germline cells (nurse cells) demonstrated the highest number of genes and UMIs likely reflecting deposition of maternal transcripts into the egg. The majority of cells produced during *Hydra* oogenesis are nurse cells that are engulfed by the single oocyte. Zymogen gland cells have a higher UMI to gene ratio, likely indicating that gland cells express a relatively small number of genes at high levels. i: cells of the interstitial lineage, id: integration doublet, ecEp: ectodermal epithelial cell, enEP: endodermal epithelial cell, mp: multiplet, nb: nematoblast, nem: differentiated nematocyte, pd: suspected phagocytosis doublet.

Metagene	Enriched Motif ID (JASPAR)	JASPAR TF Match	JASPAR TF Correlation Score	Pfam TF Match	Pfam TF Correlation Score
wg11	MA0154.3	g15090	0.414	g15090	0.414
wg13	MA0014.2	g2461	0.386	g2461	0.386
wg13	MA0779.1	g2461	0.386	g2461	0.386
wg13	MA0069.1	g2461	0.386	g2461	0.386
wg17	MA0467.1			g6880, g8768, g24114, g26815	0.403, 0.337, 0.368, 0.341
wg17	MA0234.1			g6880, g8768, g24114, g26815	0.403, 0.337, 0.368, 0.341
wg17	MA0201.1			g6880, g8768, g24114, g26815	0.403, 0.337, 0.368, 0.341
wg17	MA0609.1			g1409, g4556	0.497, 0.362
wg17	MA0844.1			g1409	0.497
wg17	MA0834.1			g1409, g4556	0.497, 0.362
wg17	MA0874.1			g6880, g8768, g24114, g26815	0.403, 0.337, 0.368, 0.341
wg17	MA0853.1			g6880, g8768, g24114, g26815	0.403, 0.337, 0.368, 0.341
wg17	MA0509.1	g17406	0.326	g17406	0.326
wg20	MA0609.1			g1409	0.364
wg20	MA0604.1			g1409	0.364
wg20	MA0844.1			g1409	0.364
wg20	MA0834.1			g1409	0.364
wg20	MA0491.1			g1409	0.364
wg25	MA0609.1			g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0834.1			g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0604.1			g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0844.1			g10604	0.373
wg25	MA0492.1	g1450	0.328	g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0148.3	g30219	0.366	g30219	0.366
wg25	MA0030.1	g30219	0.366	g30219	0.366
wg25	MA0546.1	g30219	0.366	g30219	0.366
wg25	MA0297.1	g30219	0.366	g30219	0.366
wg25	MA0488.1	g1450	0.328	g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0093.2			g31654	0.408
wg25	MA0004.1			g31654	0.408
wg25	MA0234.1	g33396	0.337	g24677, g33396	0.433, 0.337

wg25	MA0018.2			g1450, g4556, g23720	0.328, 0.323, 0.456
wg25	MA0639.1			g10604	0.373
wg25	MA0613.1	g30219	0.366	g30219	0.366
wg27	MA0014.2	g2461	0.389	g2461	0.389
wg28	MA0613.1	g25835	0.62	g25835	0.62
wg28	MA0297.1	g25835	0.62	g25835	0.62
wg28	MA0296.1	g25835	0.62	g25835	0.62
wg31	MA0798.1	g17406	0.328	g17406	0.328
wg31	MA0509.1	g17406	0.328	g17406	0.328
wg31	MA0600.2	g17406	0.328	g17406	0.328
wg31	MA0510.2			g17406	0.328
wg31	MA0799.1	g17406	0.328	g17406	0.328
wg31	MA0467.1			g6880, g8000, g18245, g24998	0.347, 0.338, 0.344, 0.371
wg31	MA0639.1			g30087	0.857
wg31	MA0853.1			g6880, g8000, g18245, g24998	0.347, 0.338, 0.344, 0.371
wg31	MA0874.1			g6880, g8000, g18245, g24998	0.347, 0.338, 0.344, 0.371
wg31	MA0854.1			g6880, g8000, g18245, g24998	0.347, 0.338, 0.344, 0.371
wg32	MA0014.2	g2461	0.69	g2461	0.69
wg32	MA0779.1	g2461	0.69	g2461	0.69
wg35	MA0297.1	g5578, g30219	0.7, 0.522	g5578, g30219	0.7, 0.522
wg35	MA0613.1	g5578, g30219	0.7, 0.522	g5578, g30219	0.7, 0.522
wg35	MA0030.1	g5578, g30219	0.7, 0.522	g5578, g30219	0.7, 0.522
wg35	MA0296.1	g5578, g30219	0.7, 0.522	g5578, g30219	0.7, 0.522
wg35	MA0546.1	g5578, g30219	0.7, 0.522	g5578, g30219	0.7, 0.522
wg35	MA0467.1			g10518	0.352
wg43	MA0259.1			g11571	0.422
wg43	MA0295.1	g25835	0.37	g25835	0.37
wg43	MA0004.1			g11571	0.422
wg43	MA0613.1	g25835	0.37	g25835	0.37
wg43	MA0297.1	g25835	0.37	g25835	0.37
wg44	MA0521.1			g26087	0.726
wg44	MA0830.1			g26087	0.726
wg44	MA0091.1			g26087	0.726
wg44	MA0522.2			g26087	0.726
wg45	MA0201.1			g28302	0.509
wg45	MA0467.1			g28302	0.509
wg45	MA0234.1			g28302	0.509
wg45	MA0295.1			g5577	0.552

wg45	MA0604.1			g28606	0.347
wg45	MA0609.1			g28606	0.347
wg45	MA0018.2			g28606	0.347
wg45	MA0608.1			g28606	0.347
wg45	MA0491.1			g28606	0.347
wg45	MA0834.1			g28606	0.347
wg45	MA0070.1			g28302	0.509
wg46	MA0014.2	g2461	0.544	g2461	0.544
wg46	MA0779.1	g2461	0.544	g2461	0.544
wg5	MA0609.1			g4556, g23720	0.316, 0.375
wg5	MA0834.1			g4556, g23720	0.316, 0.375
wg5	MA0093.2			g31654	0.45
wg5	MA0492.1			g4556, g23720	0.316, 0.375
wg5	MA0844.1			g10604	0.437
wg5	MA0604.1			g4556, g23720	0.316, 0.375
wg5	MA0526.1			g31654	0.45
wg5	MA0004.1			g31654	0.45
wg5	MA0201.1			g11765, g24677, g33396	0.309, 0.364, 0.304
wg5	MA0018.2			g4556, g23720	0.316, 0.375
wg5	MA0488.1			g4556, g23720	0.316, 0.375
wg5	MA0234.1	g33396	0.304	g11765, g24677, g33396	0.309, 0.364, 0.304
wg5	MA0692.1			g31654	0.45
wg54	MA0751.1	g27294	0.334		
wg54	MA0259.1			g8607	0.455
wg54	MA0004.1			g8607	0.455
wg54	MA0613.1	g25835	0.668	g25835	0.668
wg54	MA0297.1	g25835	0.668	g25835	0.668
wg58	MA0834.1			g4556, g23720	0.336, 0.306
wg58	MA0609.1			g4556, g23720	0.336, 0.306
wg58	MA0297.1	g30219	0.51	g30219	0.51
wg58	MA0030.1	g30219	0.51	g30219	0.51
wg58	MA0613.1	g30219	0.51	g30219	0.51
wg58	MA0844.1			g10604	0.394
wg58	MA0093.2			g31654	0.314
wg58	MA0148.3	g30219	0.51	g30219	0.51
wg58	MA0492.1			g4556, g23720	0.336, 0.306
wg58	MA0546.1	g30219	0.51	g30219	0.51
wg58	MA0604.1			g4556, g23720	0.336, 0.306

wg58	MA0852.1	g30219	0.51	g30219	0.51
wg58	MA0296.1	g30219	0.51	g30219	0.51
wg60	MA0296.1	g25835	0.344	g11764, g25835	0.404, 0.344
wg60	MA0613.1	g11764, g25835	0.404, 0.344	g11764, g25835	0.404, 0.344
wg60	MA0297.1	g11764, g25835	0.404, 0.344	g11764, g25835	0.404, 0.344
wg60	MA0259.1			g11571	0.454
wg60	MA0030.1	g11764, g25835	0.404, 0.344	g11764, g25835	0.404, 0.344
wg61	MA0804.1	g24952	0.3	g24952	0.3
wg61	MA0521.1			g8607	0.314
wg69	MA0201.1			g18245, g28302	0.354, 0.481
wg69	MA0467.1			g18245, g28302	0.354, 0.481
wg69	MA0853.1			g18245, g28302	0.354, 0.481
wg7	MA0154.3	g15090	0.353	g15090	0.353
wg71	MA0467.1			g18227	0.418
wg71	MA0201.1			g18227	0.418
wg71	MA0234.1			g18227	0.418
wg71	MA0295.1			g5577	0.439
wg76	MA0798.1	g16545	0.366	g16545	0.366
wg76	MA0510.2			g16545	0.366
wg76	MA0799.1	g16545	0.366	g16545	0.366
wg76	MA0509.1	g16545	0.366	g16545	0.366
wg83	MA0467.1			g6880, g26815	0.419, 0.326
wg83	MA0609.1			g1409, g1450, g4556	0.509, 0.314, 0.315
wg83	MA0234.1			g6880, g26815	0.419, 0.326
wg83	MA0201.1			g6880, g26815	0.419, 0.326
wg83	MA0604.1			g1409, g1450, g4556	0.509, 0.314, 0.315
wg83	MA0844.1			g1409	0.509
wg83	MA0711.1			g6880, g26815	0.419, 0.326

Table 1.5. Transcription factor candidates with putative function in cell fate specification. The JASPAR binding motif was found enriched in promoter regions of co-expressed genes within a genome metagene (MG). Co-expressed transcription factors likely to bind the identified motive were identified using two different approaches (JASPAR match, Pfam match) and are listed (see Figure 1.34). The correlation score is a measure for the expression domain overlap between a metagene and a transcription factor. A selected set of identified metagenes is presented in Figure 1.38. Metagenes can be visualized at the Broad Single-Cell Portal. For the four candidate regulators discussed in the main text we identified the corresponding transcripts (t) in the transcriptome reference by using Blast: PAX2A: g2461 - Pax-A (t9974), RFX4: g16545 – RFX4 (t30134), RX: g28302 – RX (t22218), RAX2: g18227 – HyAlx (t16456). Transcriptome IDs are presented in the main text and transcriptome data are visualized in Figure 1.5 B-D. This table contains putative regulators with a metagene correlation score > 0.3. A more extensive table (correlation score > 0.1) is available in the accompanying git repository. q-value: adjusted

enrichment p-value (Benjamini) reported by HOMER, FC: Fold-change of enrichment (calculated as percent target sequences with motif/percent background sequences with motif).

Transcript ID	Gene Name, Title	Forward	Reverse
t11061aep	APCD1	ATGCTCCTCCATCACCGATT	TGCCGTCAACAGATTCATGC
t7727aep	ARX	CTTTTAGTAGCGGGGCTGTG	TCTGGTACAGCGTCGTTAAATG
t18356aep	CHIA	AATATCGACCCGAACCAATG	AAACATGGCACCAGCAAGAC
t35005aep	CHRD	AGAAGGATGTGAACACGATGG	CCGCACTGTATGGATTTCTG
t16368aep	CO6A3	CCATGCGAGCCTGATCTTTG	ACGACATTGAGTTTCAGAGCA
t2163aep	Coma	CAGTTATTGCAACTTTGAGACGA	TGCAGTTGACAAAGAATGGGA
t2758aep	DAN	TGACAATCAAATGACTGCTAACG	TCATACACCAACTAATGCCAATG
t33064aep	desmoneme specific	AACATGGCTGCAAATGACCC	TGCCAACAGTTGTATTGCAGT
t10953aep	DKK3	ATCAGATTGCAAGCATGGTG	AGCGATGTTGGAGGTGTAGC
t2948aep	EGL4	TACCCGCACTGCTTCTGTAA	CTTGTGAAAGGAGGGTTGCC
t7356aep	EGL44	CGAAAATAAAATTGGGCGCGT	GCATTGGCGTTGTTCTCTCT
t27653aep	EHF	GGCCGTTTTCCTGAGATTCG	TCATCTGTTGCTTCGTCACT
t3974aep	ELAV2	TCAATCAGGAGGAGCTCGAC	CAGAATGAATGGCTCCGTGT
t2741aep	endoderm gradient	TGGTACATGGTCAAGGTGTTT	GTTCAACACCGCTCTTGGAG
t20960aep	HvFem-1	CACGTGCATTTCGTATTTCG	TGGAAATAGTGTTTACTGCACTG
t10354aep	HvFem-2	TGCTTTACGTGGTTTTAACACTG	TGATAAAAACATCGCTTACATGG
t12060aep	FGF1	TTGTTGTGGTTAAACTATCGCAA	CGGAGCGGTTGGTAATAAAATT
t14481aep	FGRL1	CATGAAAACGGTGCGAGTGA	TCCGGTGTCCATGGTTGTTA
t16997aep	FOX1	AGCAGCCAATAACCAGAAGC	TTTATTGTTCGACATCCATACTTTG
t12642aep	FOXL1	TCAATTTGCCGAACACAAAC	CATACGGGAAGTCGGTTACG
t15331aep	FZD8	GACACAGCTTCATCGTATCGG	AGATGCTGACGTTTGAGGGT
t29450aep	gland specific, gMGC/ZMG, basal disk	GGTTCAAGTTCAACGGCATT	AAATTGTGGAGCGTTTGATG
t38683aep	H10A	TAAGTCTCCTCCACAGTCCG	AGATGACGTGCCAACTCCTT
t11585aep	H2BL1	AGTCCAAAGAAAGGCAGTCC	GCTCAAACAAAGAAGCCGCT
t15393aep	HvSoxC	CTGAACTTACGGCGCTAAGT	TTCAGAAACTTCCGGCGTTG
t1602aep	HXB1	TTCATGTAAAACCACGCGCA	CGTTAACAACTTTGCACGCC
t27659aep	Hy-icell1	TGTTCAAATTAAACGAGATAAACTACG	AACAAGTTCTCCAATAGATTCATCG
t18862aep	HyLRR-2	TGAAAATGTACCACCCAGCG	CAGTTGTTGCGCGCTTTAAG
t4922aep	Innexin 1	AGCGAGGTTGGTTATTTTGG	ACTGGTAACCGGCTCCTTTC
t27824aep	Innexin 1A	GTATCGTGCCAGGAACAAGC	GTTGGACCACGTGATTTGACA
t23010aep	Innexin 8	TGTACCGTCTTCTGCCACTG	GATTCGCAAGGAGTTGAACC
t11055aep	LW-amide	TGTTTCGGATCCACCAAAAT	AGCTGTAGCACCCCACAAAC
t32151aep	matrilysin-like	GAGCTCGATCAATCAACCGC	TAATACGACTCACTATAGTTGGCCCAACTGTCACATTG
t27424aep	Myb	TTGCTTGAATTCGTCATTGC	AATCTCCTCGCCACATTCAC
t18095aep	Myc3	AATGATGTATGGGCAAAGTGC	AACATGAAAATCTCAAGAAAAGTCC
t13067aep	NAS14	GCGTAAGTGTAGTAGCAACTTTTCC	TAAGGTGTTCCACGCAGTTG
t5467aep	NDA-1	CGCAGATGTTTCAGGGTTCA	TTGAAATTGTGAGTTTTCCTTTG
t14976aep	NDF1	GTTCAACGAAACAAACTGCAAA	GCTAACCGTGCATTGAGGAA
t28450aep	neuro cluster4 marker	GCTGTAAATCTAGCACCTCTAAATCA	ATAGCTTCACTGGTTTAAACAAAAC
t6329aep	neuro cluster7 marker	CACATCACAATTAGAAAACTGTTCG	TTAAAAGTGCATTTGGGTTTGA
t28441aep	NR2E1	ATGTGATGGATGTCGTGGGT	CAGGTGTTTGTGGGTGAGTG
t15240aep	pyk3-like	CGAACTGTCAAACTGCAAGC	CCAAGGATTGAGTTGCAAGG
t25612aep	RE114	TGGAAAATAGAGAAGTATGGACG	TGCAGGAAAATTTGGGTCAA
t33899aep	RFamide preprohormone E	CGTTGTTTTCGCTTAGTTAGCTG	TCTTTTCCGAACCGAGGTAG
t22135aep	RSGI5	CGGACAAAACGGGACTGTAT	ATCGCTCAAACTCAGGGATG
t17610aep	S17B1	GGAACTTTTGCCTGCCTCAA	GCGTTACTGATGATTGCTTGC
t13827aep	spermatogenesis	CAACTTCGCGTCGCTAATGA	ATACCCAGCACAGCAACAAC
t19014aep	spermatogenesis	TGAAGAGCGTGTATCCTCAGA	TTCGTGCGCCTTATTGTGAC
t19036aep	SRFP3	TCTGCTGCTATCATCTGCTG	TCAGCTTTGTTTTCCACACC
t10854aep	stenotele specific	TGGATTAGACGAGTGGACCG	GTCATAGTGCGCCATTCTCG
t2754aep	SYCP2	AGTTTGGAAATTTGCAGCACT	ACTGCAATGATGGATAGGTAGC
t33926aep	TEAD1	GCATGACTGGTCTCGAAGTA	AGGTGTTGCTGGGTGATGAT
t11591aep	ZN333	AGTCGATCTATACCGTGGCA	ACTTTCGACACTGCCAACAC
t1609aep		TGGTTGAACACTGAGCAGGA	AAAATATCCACATGACGGTTCG
t17610aep	S17B1	GGAACTTTTGCCTGCCTCAA	GCGTTACTGATGATTGCTTGC

Table 1.6. Primers. Primer pairs used for RNA probe generation to be used in RNA in situ hybridization.

Transcript ID	*Hydra* gene	Accession
t31074aep	arminin1a	GU256274.1
t262aep	Bmp2/4	AB823952
t18735aep	budhead	AY263364.1
t15465aep	Cerberus-like 4	XP_002162057.1
t10853aep	CnASH	U36275
t25396aep	CnNK-2	AF012538.1
t11407aep	Cnnos1	AB037080
t1163aep	Cnot	AJ252184.1
t17278aep	CnOtx	AF114441.1
t18095aep	HvMyc3	LN868213.1
t15393aep	HvSoxC	JQ994230.1
t474aep	HvTSP	XM_012702849.1
t16456aep	HyAlx	AF295531.1
t20768aep	Hybra1	AF105065, AY366371
t29725aep	Hybra2	AY366372
t22117aep	HyDkk-3	AY332609.1
t14102aep	HyDkk1/2/4_C	NM_001280836.1
t8678aep	HyDkk1/2/4_A	NM_001280833.1
t18862	HyLRR2	NM_001280922.1
t1679aep	Hym-176A	AB018544.1
t17992aep	Hym-176B	Hm02822
t6329aep	Hym-176C	Hm00388
t12588aep	Hym-176D	Hmp15428
t33899aep	Hym-176E	Hmp21432
t12874aep	Hym-355	AB025945.1
t10549aep	Hym301	AB106883.1
t12596aep	HyTSR1	AM182484.3
t34367aep	Hywi	AB840994.1
t15597aep	HyWnt1	AB426122.1

Transcript ID	*Hydra* gene	Accession
t14194aep	HyWnt3	AF272673.1
t21554aep	HyWnt5a	AM263447.1
t28874aep	HyWnt7	AB426121.1
t23521aep	HyWnt8	NM_001309735.1
t10028aep	HyWnt9/10c	AB426119.1
t4922aep	Innexin 1	DQ372935.1
t27824aep	Innexin 1A	XP_002165135.1
t8891aep	Innexin 2	XP_002160488
t23010aep	Innexin 8	XP_002164718.1
t4498aep	ks1	X78596.1
t11055aep	LW-amide	U53444.1
t5467aep	NDA-1	XM_002162825.3
t5467aep	NDA-1	XM 002162825
t23176aep	nematocilin A	BAG48261
t13480aep	nematogalectin B	AB583745
t15237aep	nowa-1	NM 001287375.1
t9974aep	Pax-A	AAB58290.1
t11117aep	periculin1a	FJ517724.1
t5275aep	Pitx	XM 002164950.3
t16043aep	PPOD1	DQ073555.1
t20256aep	prdl-a	Y15515.1
t3809aep	RFamide preprohormone A	Y11678
t2059aep	RFamide preprohormone B	Y11679
t25706aep	RFamide preprohormone C	Y11680
t16657aep	RFamide preprohormone D	BP507974
t2965aep	SYCP1	JQ906935.1
t2390aep	SYCP3	JQ906933.1

Table 1.7. Transcript annotations of published *Hydra* genes used in this study.

Transcript ID	Name/Description	PFAM (eval 1e-6)	SP e-val	Swiss-Prot ID	Swiss-Prot Name	NR ID	NR e-val	NR Title
t1609aep		ShK(PF01549.23)				XP_012565484	8.00E-135	PREDICTED: uncharacterized protein LOC100207073 isoform X1 [Hydra vulgaris].
t11061aep	APCD1	APCDDC (PF14921)	1.85E-24	Q5R2I8|APCD1_CHICK	Protein APCDD1	XP_002165888.1	0	PREDICTED: protein APCDD1- like [Hydra vulgaris]
t7727aep	ARX	Homeobox (PF00046)	2.94E-25	Q96QS3|ARX_HUMAN	Homeobox protein ARX	XP_002163599.1	1.15E-146	PREDICTED: homeobox protein ARX-like [Hydra vulgaris]
t18356aep	CHIA	Glyco_hydro_18 (PF00704)	2.65E-111	Q91XA9|CHIA_MOUSE	Acidic mammalian chitinase	XP_002163322.2	0	PREDICTED: chitinase-3-like protein 1 [Hydra vulgaris]
t35005aep	CHRD	VWC (PF00093)	1.57E-15	Q91713|CHRD_XENLA	Chordin	XP_002158106.2	0	PREDICTED: chordin-like protein 1 isoform X1 [Hydra vulgaris]
t16368aep	CO6A3	Kunitz_BPTI (PF00014.22), IGFBP(PF00219.17)	1.00E-12	P15989|CO6A3_CHICK	Collagen alpha-3(VI) chain	XP_002156558	1.00E-84	PREDICTED: venom protein 302-like [Hydra vulgaris].
t2163aep	COMA	Phospholip_A2_3 (PF09056.10)	2.00E-09	Q9TWL9|COMA_CONMA	Conodipine- M alpha chain	XP_002156354	5.00E-121	PREDICTED: uncharacterized protein LOC100205674 [Hydra vulgaris].
t2758aep	DAN domain	DAN domain						
t33064aep	desmoneme specific					XP_002167712.3	8.96E-129	PREDICTED: uncharacterized protein LOC100209667 [Hydra vulgaris]
t34731aep	desmoneme specific	Collagen (PF01391)						
t10953aep	DKK3	Dickkopf_N (PF04706)	1.74E-09	Q9QUN9|DKK3_MOUSE	Dickkopf- related protein 3	XP_012566156.1	2.12E-81	PREDICTED: dickkopf-related protein 3-like [Hydra vulgaris]
t21290aep	DMC1	Rad51 (PF08423), RecA (PF00154)	0	Q14565|DMC1_HUMAN	protein DMC1/LIM15 homolog	NP_001274726.1	0	DMC1 homologue CnDMC1 [Hydra vulgaris]
t2948aep	EGL4	Pkinase(PF00069.24),cNMP_binding(PF00027.28),Pkinase_Tyr(PF07714.16)	0	O76360|EGL4_CAEEL	cGMP- dependent protein kinase egl-4	XP_002154974	0	PREDICTED: cGMP-dependent protein kinase 1-like [Hydra vulgaris]
t7356aep	EGL44	TEA (PF01285)	4.08E-31	Q19849|EGL44_CAEEL	Transcription enhancer factor-like protein egl- 44			
t27653aep	EHF	Ets (PF00178)	9.84E-20	Q32LN0|EHF_BOVIN	ETS homologous factor	XP_012555395.1	0	PREDICTED: uncharacterized protein LOC105843939 [Hydra vulgaris]
t3974aep	ELAV2	RRM_1 (PF00076)	4.99E-94	Q28GD4|ELAV2_XENTR	ELAV-like protein 2	XP_002156791.3	0	PREDICTED: ELAV- like protein 3 [Hydra vulgaris]
t2741aep	endo_gradient					WP_051398353.1	4.59E-107	hypothetical protein [Runella limosa]
t22116aep	ETV1	Ets (PF00178),ETS_PEA3_N (PF04621)	1.55E-52	P41164|ETV1_MOUSE	ETS translocation variant 1	XP_002166140.2	0	PREDICTED: ETS translocation variant 1-like isoform X2 [Hydra vulgaris]
t12060aep	FGF1	FGF(PF00167.17), Fascin(PF06268.12)	1.00E-07	Q6I6M7|FGF1_CYNPY	Fibroblast growth factor 1	AND74488	2.00E-112	FGF-1 [Hydra vulgaris]

t14481aep	FGRL1	Ig_3(PF13927.5), I-set(PF07679.15), Ig_2(PF13895.5), ig(PF00047.24), V-set(PF07686.16)	3.00E-17	Q7T2H2|FGRL1_CHICK	Fibroblast growth factor receptor-like 1	XP_002163676	0	PREDICTED: fibroblast growth factor receptor- like isoform X1 [Hydra vulgaris]
t16997aep	FOX1	RRM_1 (PF00076)	4.18E-27	Q10572|FOX1_CAEEL	Sex determinatio n protein fox-1	XP_002155955.3	2.34E-174	PREDICTED: sex determination protein fox-1-like [Hydra vulgaris]
t12642aep	FOXL1	Forkhead (PF00250)	9.32E-17	Q64731|FOXL1_MOUSE	Forkhead box protein L1	XP_002166935.1	0	PREDICTED: forkhead box protein C1-like [Hydra vulgaris]
t15331aep	FZD8	Fz (PF01392)	4.74E-31	Q498S8|FZD8_RAT	Frizzled-8	XP_002163691.2	0	PREDICTED: frizzled-8-like [Hydra vulgaris]
t29450aep	Gland specific (gMGC/ZMG, basal disk)					XP_002166617.1	2.88E-174	PREDICTED: multiple epidermal growth factor-like domains protein 11 [Hydra vulgaris]
t38683aep	H10A	Linker_histone (PF00538)	1.54E-18	P22845|H10A_XENLA	Histone H1.0-A	XP_002156243.2	8.05E-45	PREDICTED: histone H1.0-B-like [Hydra vulgaris]
t11585aep	H2BL1	Histone (PF00125)	6.16E-50	P07794|H2BL1_PSAMI	Late histone H2B.2.1	XP_002162131.1	1.69E-67	PREDICTED: late histone H2B.L4- like [Hydra vulgaris]
t20960aep	HvFem-1 (Hydra female 1)					MK648246		
t10354aep	HvFem-2 (Hydra female 2)					MK648245		
t1602aep	*HXB1*	Homeobox (PF00046.28)	3.00E-16	P81192|HXA4_LINSA	Homeobox protein Hox- A4	NP_001296687	2.00E-165	homeobox protein Hox-D10-like [Hydra vulgaris]
t27659aep	*Hy-icel1*					MK648244		
t32151aep	matrilysin-like	Peptidase_M10 (PF00413), PG_binding_1 (PF01471), ShK (PF01549)	1.29E-42	P22757|HE_PARLI	Hatching enzyme	XP_002164594.2	0	PREDICTED: matrilysin-like [Hydra vulgaris]
t7059aep	MUC2	VWD (PF00094),C8(PF08742),TIL (PF01826),VWC(PF00093)	1.71E-157	Q62635|MUC2_RAT	Mucin-2	XP_004206639.2	0	PREDICTED: mucin-5AC-like [Hydra vulgaris]
t27424aep	*Myb*	Myb_DNA-bind_4(PF13837)				MK648243	0	PREDICTED: uncharacterized protein LOC101237947 [Hydra vulgaris]
t13067aep	NAS14	Astacin(PF01400)	4.54E-41	Q19269|NAS14_CAEEL	Zinc metalloprote inase nas-14	XP_002158506.2	0	PREDICTED: astacin-like metalloprotease toxin 5 [Hydra vulgaris]
t7084aep	NAS15	Astacin(PF01400)	1.28E-25	P55115|NAS15_CAEEL	Zinc metalloprote inase nas-15	XP_012560695.1	1.74E-130	PREDICTED: zinc metalloproteinase nas-15-like, partial [Hydra vulgaris]
t14976aep	NDF1	HLH(PF00010)	1.51E-08	O42202|NDF1_DANRE	Neurogenic differentiatio n factor			
t28450aep	neuro cluster4 marker							
t6329aep	neuro cluster7 marker							

t33301aep	marker Alpha- LTX-Lhe1a-like					XP_012558950.1	1.02E-56	PREDICTED: uncharacterized protein LOC105845581 [Hydra vulgaris]
t28441aep	NR2E1	Hormone_recep(PF00104), zf-C4(PF00105)	9.48E-77	Q9Y466|NR2E1_HUMAN		XP_002154441.1	0	PREDICTED: nuclear receptor subfamily 2 group E member 1-like [Hydra vulgaris]
t16434aep	ODF3A	SHIPPO-rpt(PF07004)	3.81E-76	Q8AVY1|ODF3A_XENLA	Outer dense fiber protein 3	XP_002157290.2	4.02E-178	PREDICTED: outer dense fiber protein 3-like [Hydra vulgaris]
t10355aep	PCNA	PCNA_C(PF02747),PCNA_N (PF00705)	6.41E-103	P61258|PCNA_MACFA	Proliferating cell nuclear antigen	XP_002164838.1	1.73E-176	PREDICTED: proliferating cell nuclear antigen- like [Hydra vulgaris]
t15240aep	pyk3-like					XP_002159399.1	0	PREDICTED: dual specificity protein kinase pyk3-like [Hydra vulgaris]
t25612aep	RE114	REC114-like(PF15165)				XP_002170240.2	3.15E-80	PREDICTED: pol- RFamide neuropeptides-like [Hydra vulgaris]
t33899aep	RFamide preprohormone E							
t5694aep	RFX6	RFX_DNA_binding(PF02257)	4.90E-159	Q5RJA1|RFX6_DANRE	DNA-binding protein RFX6	XP_012560423.1	0	PREDICTED: uncharacterized protein LOC101239106 [Hydra vulgaris]
t22135aep	RSGI5	AbfB(PF05270)	2.27E-06	A3DEX6|RSGI5_CLOTH	Anti-sigma-I factor RsgI5	XP_002159690.1	0	PREDICTED: uncharacterized protein LOC100202073 [Hydra vulgaris]
t22218aep	RX	Homeobox(PF00046)	4.64E-15	O35602|RX_MOUSE	RX_MOUSE Retinal homeobox protein Rx	XP_012566844.1	4.97E-128	PREDICTED: short stature homeobox protein-like [Hydra vulgaris]
t17610aep	S17B1	HMG_box(PF00505), HMG_box_2(PF09011)	1.29E-22	O42601|S17B1_XENLA	Transcription factor Sox- 17 beta.1	XP_002156236.1	0	PREDICTED: sex- determining region Y protein- like [Hydra vulgaris]
t19036aep	SFRP3	NTR(PF01759), Fz(PF01392)	8.37E-25	Q95117|SFRP3_BOVIN	Secreted frizzled- related protein 3	XP_012556435.1	0	PREDICTED: secreted frizzled- related protein 3- like [Hydra vulgaris]
t13827aep	spermatogenesis specific					XP_002156844.3	1.12E-06	PREDICTED: uncharacterized protein LOC100212533 [Hydra vulgaris]
t19014aep	spermatogenesis specific							
t10854aep	stenotele specific	DUF3472(PF11958),DUF5077(PF16871)				XP_002157558.3	0	PREDICTED: nematoblast specific protein [Hydra vulgaris]
t35089aep	stenotele specific					NP_001267858.1	1.26E-25	nematoblast- specific protein precursor [Hydra vulgaris]
t2754aep	SYCE2		2.66E-08	Q505B8|SYCE2_MOUSE	Synaptonemal complex central element protein 2	NP_001296666.1	8.24E-106	synaptonemal complex central element protein 2 [Hydra vulgaris]
t33926aep	TEAD1	TEA(PF01285)	3.02E-25	P28347|TEAD1_HUMAN	Transcription al enhancer factor TEF-1	XP_012553729.1	0	PREDICTED: uncharacterized protein LOC105843218 [Hydra vulgaris]
t11591aep	ZN333	zf-H2C2_2(PF13465), zf-C2H2(PF00096), zf-C2H2_4(PF13894), zf-C2H2_11(PF16622), zf-H2C2_5(PF13909), HNF-1_N(PF04814)	6.78E-39	Q96JL9|ZN333_HUMAN	Zinc finger protein 333	XP_002157355.1	0	PREDICTED: zinc finger protein 26- like [Hydra vulgaris]

Table 1.8. Transcript annotations for uncharacterized genes discussed in this study. The Swiss-Prot and NCBI nr databases (E-value ≤ 1e-5) were searched using blastx. The Pfam v24.0 database was searched to identify protein domains (E-value ≤ 1e-6). InterPro (*https://www.ebi.ac.uk/interpro/)* was searched in cases were no results were obtained in blastx and Pfam searches.

CHAPTER 2. Differentiation trajectories of the *Hydra* nervous system reveal transcriptional regulators of neuronal fate

This chapter was originally submitted as pre-print on *BioRxiv:*

Primack AS, JF Cazet, H Morris Little, S Mühlbauer, BD Cox, CN David, JA Farrell, CE Juliano. 2023. Differentiation trajectories of the Hydra nervous system reveal transcriptional regulators of neuronal fate. bioRxiv *2023.03.15.531610*

ABSTRACT

The small freshwater cnidarian polyp *Hydra vulgaris* uses adult stem cells (interstitial stem cells) to continually replace neurons throughout its life. This feature, combined with the ability to image the entire nervous system (Badhiwala et al., 2021; Dupre and Yuste, 2017) and availability of gene knockdown techniques, makes *Hydra* a tractable model for studying nervous system development and regeneration at the whole-organism level. In this study, we use single-cell RNA sequencing and trajectory inference to provide a comprehensive molecular description of the adult nervous system. This includes the most detailed transcriptional characterization of the adult *Hydra* nervous system to date. We identified eleven unique neuron subtypes together with the transcriptional changes that occur as the interstitial stem cells differentiate into each subtype. Towards the goal of building gene regulatory networks to describe *Hydra* neuron differentiation, we identified 48 transcription factors expressed specifically in the *Hydra* nervous system, including many that are conserved regulators of neurogenesis in bilaterians. We also performed ATAC-seq on sorted neurons to uncover previously unidentified putative regulatory regions near neuron-specific genes. Finally, we provide evidence to support the existence of transdifferentiation between mature neuron subtypes and we identify previously unknown transition states in these pathways. All together, we provide a comprehensive transcriptional description of an entire adult nervous system, including differentiation and transdifferentiation

pathways, which provides a significant advance towards understanding mechanisms that underlie nervous system regeneration.

INTRODUCTION

Understanding the fundamental principles of nervous system regeneration is key to developing treatments for traumatic brain injury, spinal cord injury, and neurodegenerative diseases. Most well-established neurobiology research organisms are not able to replace adult neurons, making their utility for investigating adult nervous system regeneration limited. By contrast, the small freshwater cnidarian polyp *Hydra vulgaris* (Figure 2.1 A) has active adult multipotent stem cells that renew its entire nervous system throughout its life (David and Gierer, 1974). In addition, the relative simplicity of *Hydra* allows for the study of adult nervous system development and regeneration at the level of the entire nervous system. *Hydra* has a relatively simple and well-understood body that is composed of two epithelial monolayers, an outer ectodermal layer and an inner endodermal layer. The nervous system is made up of approximately 3,000-5,000 neurons (about 3-5% of all cells in the body) arranged into two separate nerve nets that are embedded in the two epithelial layers (Keramidioti et al., 2023). *Hydra* neurons are part of the interstitial cell lineage (David, 2012; David and Murphy, 1977) and are supported by multipotent adult interstitial stem cells (ISCs) that are found in the interstices of the ectodermal epithelial body column cells (Figure 2.1 B). Due to passive tissue displacement towards the extremities that result in perpetual cell loss (Figure 2.1 A arrows), the ISCs continuously replace neurons (Richard D Campbell, 1967; David and Gierer, 1974; Hager and David, 1997) such that all neural differentiation pathways are active in the adult polyp and are therefore accessible for

experimentation. The ISCs also enable regeneration of the entire nervous system following injury.

In addition to exhibiting continual, widespread neuronal regeneration, *Hydra* are amenable to genetic manipulations, such as gene knock-down through the expression of RNA hairpins (Juliano et al., 2014c; Lohmann et al., 1999) or the electroporation of siRNAs (Lohmann et al., 1999; Vogg et al., 2022). Additionally, it is straightforward to create stable transgenic *Hydra* lines using cell-type-specific promoters
(Dupre and Yuste, 2017; Klimovich et al., 2019; Siebert et al., 2019; Wittlieb et al., 2006). Taken together, these attributes make *Hydra* an accessible model for neurobiology, enabling us to test the developmental regulators of neuron differentiation for an entire nervous system. Finally, *Hydra*'s position on the phylogenetic tree as a member of Cnidaria, the sister group to bilaterians, enables researchers to explore the evolution of the molecular mechanisms underlying nervous system development and regeneration.

Decades of research provide a basic framework for *Hydra* neurobiology, including the source of new neurons (David, 2012), the relative distribution of the nervous system along the oral-aboral axis (Bode et al., 1973), and the rates of neurogenesis in an uninjured *Hydra* (David and Gierer, 1974). As part of our previous study creating a whole animal single cell expression map, we transcriptionally profiled *Hydra* neurons and determined their spatial location (Siebert et al., 2019). However due to the relatively small number of neurons sequenced (~3,500), it remained unclear whether we successfully profiled all populations of differentiated neurons. Additionally, in our original study we did not capture enough intermediate states of neuronal differentiation or

transdifferentiation to properly resolve the transcriptional changes that occur as neurons are produced during homeostatic maintenance.

We resolved these issues by increasing our data set by 10-fold to ~35,000 single-cell transcriptomes of neurons and neural progenitors. From these data, we identified eleven neuron subtypes in *Hydra*: three in the endodermal nerve net and eight in the ectodermal nerve net. These data are largely consistent with our previously published spatial molecular map of the *Hydra* nervous system (Siebert et al., 2019). We used our data to build differentiation trajectories describing the transcriptional changes that underlie the homeostatic differentiation of all eleven neuron subtypes in *Hydra*. We also provide new evidence to support widespread transdifferentiation between neuron subtypes in the *Hydra* nervous system. We performed neuron-enriched Assay for Transposase-Accessible Chromatin (ATAC-seq) (Buenrostro et al., 2013; Corces et al., 2017) to identify the chromatin states of *Hydra* neurons, and we identified the transcription factors that are expressed during the differentiation of all *Hydra* neuron subtypes. Our data describe the differentiation of a complete nervous system and identify putative key regulators of each neuronal subtype for future testing. Furthermore, our data lay the groundwork for studies investigating how these pathways are activated to regenerate the *Hydra* nervous system in response to injury.

RESULTS

The *Hydra* nervous system is composed of eleven transcriptionally distinct neuron subtypes

In our previous study, we used Drop-seq (Evan Z. Macosko et al., 2015) to build a single cell atlas of the adult *Hydra* polyp, which included approximately 3,500 single-cell transcriptomes of differentiated neurons and cells undergoing neurogenesis (Siebert et al., 2019). In this present study, we first aimed to increase the number of neural single-cell transcriptomes to reveal any

molecular diversity in the *Hydra* neuron repertoire that we may have previously missed due to the relatively low number of cells sequenced. To this end, we used scRNA-seq (10x Genomics) to increase the number of sequenced neurons and neural progenitors to ~35,000 single-cell transcriptomes.

To enrich for neurons and neural progenitors, we used Fluorescent Activated Cell Sorting (FACS) to collect cells from two different transgenic lines: (1) *Tg(actin1:GFP)*$^{rs3\text{-}in}$, in which GFP is expressed in all differentiated neurons, neural progenitors, and ISCs (Keramidioti et al., 2023), and (2) *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$, which was created for this study and in which mNeonGreen is predicted to be expressed in all differentiated neurons (Figures 2.6 and 2.7; Table 2.1). We combined our new data with the neuronal single-cell transcriptomes we previously collected (Siebert et al., 2019) for downstream processing (Table 2.2). For all libraries, the sequencing reads were mapped to the *Hydra vulgaris* strain AEP gene models (Cazet et al., 2023) and processed following standard procedures (Figure 2.8) (Evan Z. Macosko et al., 2015). After filtering, we recovered 35,071 single-cell neural transcriptomes with a detected median of 1371.5 genes and 2887.5 UMIs per cell (Table 2.2). This is in comparison to the detected median of 563.5 genes and 1082 UMIs per cell in the neuron-enriched libraries collected by Drop-seq in our previous study (Siebert et al., 2019).

To identify distinct neuron subtypes in our data set, we used Seurat to perform Louvain clustering and visualized the results using Uniform Manifold Approximation and Projection (UMAP) (Figure 2.1 C) (Hao et al., 2021a; McInnes et al., 2018; Satija et al., 2015; Stuart et al., 2019). We recovered and annotated 11 neural clusters using marker genes identified previously

(Siebert et al., 2019): three neuron subtypes that reside in the endodermal nerve net (en1, en2, and en3) and eight neuron subtypes that reside in the ectodermal nerve net (ec1A, ec1B, ec2, ec3A, ec3B, ec3C, ec4, and ec5) (Figure 2.1 C-D, Figure 2.9, Table 2.3). We also recovered three clusters that express marker genes for multiple subtypes that we hypothesize are in a transition state (td1, td2, td3) (discussed further below). We also annotated stem cell and neural progenitor cell clusters using markers from our previous study (Siebert et al., 2019).

This analysis aligned with our previously published map (Siebert et al., 2019) with one small discrepancy regarding the previously annotated ec4A and ec4B neuron subtypes. Although ec4A and ec4B clustered separately in our previously published map, it is unclear if this distinction is justified in our current analysis and we therefore have collapsed this into a single ec4 population. Future work should more carefully define this population on morphological and functional levels. Finally, we performed non-negative matrix factorization (NMF) (Kotliar et al., 2019) to identify groups of co-expressed genes ("metagenes"), and we recovered at least one metagene for each of the eleven neuron subtypes, as well as for the three newly annotated transdifferentiation cell states (Figure 2.10).

In summary, our new scRNA-seq data set enabled us to achieve deeper sequencing, significantly increase the number of neural progenitor transcriptomes, and profile rarer populations of cells potentially undergoing transdifferentiation. Below, we use these data to profile the transcription factors expressed in the entire nervous system and resolve stem cell differentiation and transdifferentiation trajectories.

Comprehensive identification of transcription factors selectively expressed in the *Hydra* nervous system

Our work makes *Hydra* one of the few adult organisms in which the entire nervous system has been transcriptionally defined. Given the relative simplicity of only 11 neuron subtypes, *Hydra* provides the opportunity to build gene regulatory networks (GRNs) that describe the differentiation of all neurons in an adult nervous system. This requires both identifying the transcription factors (TFs) expressed during nervous system development as well as determining the cis-regulatory elements (CREs) that are bound by those TFs. To provide starting points for this long-term goal, we used our scRNA-seq data sets to identify the TFs expressed in the *Hydra* nervous system, but not in other cell types.

In our previous work, we identified 811 putative transcription factors (TFs) (i.e., genes with predicted DNA binding domains) with detectable expression in *Hydra* polyps (Cazet et al., 2023). Here, we find that 48 of these TFs are selectively expressed in the *Hydra* nervous system (Figures 2.2 and 2.11). Of these 48 TFs, four are expressed in neural progenitors (*myc3, myb, rfx5/7, hmgb3*); five are expressed in different endodermal neurons (neurog1/2/3*, G004963, duxa/b, klf3, ptf1a*); 12 are expressed in different ectodermal neurons (*smad4, sox3, foxl2, zic1/2/3/4/5, litaf, G018876, noto, sox2, arx, bhlha15, gata3, ndf1, G010046*) and the remainder are either co-expressed in ectodermal and endodermal neurons or are pan-neuronally expressed.

As mentioned above, building GRNs to describe nervous system development in *Hydra* will also require determining CREs that are bound by the neural TFs. As a first step towards this goal, we performed ATAC-seq on FACS-sorted *Hydra* neurons from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ and

Tg(tba1c:mNeonGreen)$^{cj1\text{-}gt}$ transgenic lines to reveal the cis-regulatory landscape of *Hydra* neurons (Table 2.4). The data produced were high quality (based on ENCODE standards, (encodeproject.org/atac-seq) (Landt et al., 2012) and replicates exhibited high reproducibility (Figure 2.12). Data quality was similar to three previously published whole *Hydra vulgaris* strain AEP ATAC-seq data sets (Cazet et al., 2023), suggesting that FACS sorting did lower the quality of the nuclei collected.

We found that peaks from the neuronal ATAC-seq libraries collected from both transgenic lines had significantly increased accessibility near neural genes as compared to whole animal ATAC-seq data ($p < 0.001$ for each line; Figure 2.3 A-B). Peak accessibility near some neuronal subtype-specific genes varied between samples generated from different transgenic lines but reflected the cell types collected from those transgenic lines (Figure 2.13). Notably, these data identified accessible regions near neuronal genes that were not detected in our previously collected whole-animal data (Cazet et al., 2023), which demonstrates the utility of these data for identifying the regulatory regions that drive neuronal expression (Figure 2.3 C-H).

Differentiation trajectories reveal transition states during development of an entire nervous system

Having defined the repertoire of neuron subtypes in *Hydra*, as well as having identified the transcription factors selectively expressed in the nervous system, we next aimed to determine the ordering of transcription factor expression over developmental time. The *Hydra* nervous system undergoes continual renewal with complete nervous system turnover approximately every three weeks (Bode et al., 1988). Therefore, when performing scRNA-seq on the adult *Hydra*, we

expect to profile stem cells, differentiated cells, and cells in various states of the differentiation process (Siebert et al., 2019). Thus, we used our scRNA-seq data to build differentiation trajectories for each of the 11 neuron subtypes using URD (Figure 2.4 A), which is a diffusion-based approach to generate developmental trajectories in the form of branching trees (Farrell et al., 2018). URD requires that we define both the root and the tips of the tree. In this case, the root is ISCs, which were defined by a specific marker (*G002332*) (Siebert et al., 2019), and the tips were the neuron subtypes as defined by their expression of specific markers (Figure 2.9) and lack of progenitor markers.

Our previous work showed that as ISCs begin the process of neurogenesis, they express the transcription factors *myc3* (*G003730*) and *myb* (*G020130*); the expression of these genes is lost as neurons complete differentiation (Siebert et al., 2019). Therefore, we used the expression of *myc3* and *myb* to identify the neural progenitors in our trajectory analysis and found that two populations of cells expressing both *myc3* and *myb* initially split from the ISCs. One of these progenitor populations (Figure 2.4 A, segment 3) gives rise to all endodermal neurons (en1, en2, en3) and the other (Figure 2.44 A, segment 2) gives rise to most ectodermal neurons (ec1A, ec1B, ec2, ec3A, ec3B, ec3C, ec4). Intriguingly, the ec5 ectodermal neurons (Figure 2.4 A, segment 19) did not connect to either of these populations and rather appeared to differentiate directly from ISCs. This suggests that these neurons may either differentiate directly, which would be surprising (David, 2012), or that there is an additional progenitor population or transition state that was not captured in our data set. Regardless, our data reveal that most endodermal and ectodermal neurons likely arise from two distinct progenitor populations. As

differentiation proceeds, neural progenitors become more restricted in fate potential, ultimately giving rise to the eleven neuron subtypes.

The differentiation trajectory also allows us to track TF dynamics over developmental time. To visualize this, we plotted the temporal expression of genes, with a focus on TFs, that are expressed during the differentiation of each neuron subtype. For example, we tracked the expression of genes *bhlha15 (G021353)*, *gata3 (G022640), hym355 (G004115)*, and ec3A marker *G021930* over the course of ec3A differentiation (Figure 2.4 A, segments 1, 2, 4, 11, Figure 2.4 B). To validate the transition states that our trajectory predicts for ec3A differentiation, we used double fluorescent *in situ* hybridization (FISH), focusing on pairs of genes with overlapping expression domains during three predicted transition states of ec3A differentiation (Figure 2.4 C-U). For each pair of genes (Figure 2.4 C-G *bhlha15* + *gata3*, Figure 2.4 H-L *gata3* + *hym355*, Figure 2.4 M-Q *hym355* + *G021930*), we found cells that individually expressed the genes as well as cells that co-expressed the pair of genes, as predicted by our differentiation trajectory. We also found the majority of stained neurons co-expressing *hym355* + *G021930* in the aboral end (Figure 2.4 M, R-U), which is where we expect to find ec3A neurons. This confirms that our trajectory analysis identified the dynamics of gene expression during the specification and differentiation of individual neuronal subtypes in *Hydra*.

Transcriptional evidence of neuron transdifferentiation in *Hydra*

Although our trajectory analysis identified the dynamics of gene expression when new neurons are born from ISCs, our data also revealed strong evidence of transdifferentiation between differentiated neuron subtypes. Due to the tissue dynamics in *Hydra* (Figure 2.1 A,D), neurons

are continually displaced towards the extremities, which has led to speculation that neurons undergo transdifferentiation to accommodate their new position (Bode, 1992). Some studies have documented neuron transdifferentiation in *Hydra*, but these studies were done in animals that lacked ISCs (Bode et al., 1988; Bode, 1992; Koizumi and Bode, 1991, 1986). *Hydra* lacking ISCs would be unable to make neurons *de novo* and may thus be forced to activate normally unused developmental pathways. We therefore looked for evidence of neuronal transdifferentiation in our data set to determine if this was a common phenomenon in unmanipulated *Hydra*. In addition to the eleven previously identified neuron subtypes, we found three additional clusters [annotated as transdifferentiation (td) 1, 2, and 3] (Figure 2.1 C). We hypothesize that these clusters represent neurons in the process of transdifferentiation because they express markers of multiple previously identified neuron subtypes. Based on the co-expression of ec1A and ec1B markers (Figure 2.5 A), we hypothesize that transdifferentiation clusters 1 and 3 represent ec1A neurons from the body column transdifferentiating into ec1B neurons as they are displaced toward the oral end. Based on the co-expression of ec1A and ec5 markers (Figure 2.5 A), we hypothesize that the transdifferentiation cluster 2 represents ec1A neurons from the body column transdifferentiating into ec5 peduncle neurons as they are displaced toward the aboral end.

To uncover possible transdifferentiation events**,** we visualized cellular relationships in a structure agnostic manner as a complementary approach to URD using a Force Directed Layout (FDL) (Farrell et al., 2018; Fruchterman and Reingold, 1991). We isolated groups of neurons that we hypothesized undergo transdifferentiation with each other and generated Force Directed Layouts (FDLs) based on weighted k-nearest neighbor networks, where linkages were stronger when cells were more transcriptionally similar (Fruchterman and Reingold, 1991; Jacomy et al., 2014)

(Figure 2.5 B-D). Although not demonstrated directly in the data, we inferred the directionality of transdifferentiation events in this analysis based on the known direction of cell movements (from the body column toward extremities) and the anatomical location of individual neuronal populations (Figure 2.1 D, Figure 2.5 B-D). For the first group of cells (ec1A, ec1B, ec5, td1, td2, td3), we found that ec1A connects to ec1B through two intermediate transdifferentiation populations (td1 and td3) (Figure 2.5 B). Additionally, we found that ec1A weakly connects to ec5 both directly and possibly through an intermediate state of td2, although the relationship of td2 to ec1A and ec5 is less clear. This provides further evidence that body column ec1A cells transdifferentiate into head ec1B cells, and possibly peduncle ec5 neurons. For the second group of cells, (ec3A, ec3B, ec3C, ec3_precursors), ec3A and ec3C appear to be derived from ec3 precursors in addition to undergoing transdifferentiation from body column ec3B neurons as they move into the extremities (Figure 2.5 C). As a control, we tested the interactions of several neuron subtypes with little to no predicted transcriptional similarity (Figure 2.5 D). As expected, there were no strong connections between the control clusters.

DISCUSSION

The number of research organisms used to study nervous system development and function has substantially increased in recent years due to advances in sequencing technology (Albertin et al., 2022; Fincher et al., 2018; Hulett et al., 2022, 2020; Orvis et al., 2022; Sachkova et al., 2021; Sebé-Pedrós et al., 2018). This has enabled researchers to characterize the transcriptional diversity in neurons from a wide range of organisms. However, very few comprehensive transcriptional descriptions of adult nervous systems currently exist. Here, we provide a transcriptional analysis of the entire *Hydra vulgaris* nervous system. Although *Hydra* shares many of the same experimental advantages as well-established invertebrate models, such as

small size, optical transparency, and ability to test gene function, it is also able to regenerate its entire nervous system from adult stem cells (Bode et al., 1988; Dupre and Yuste, 2017). We built a molecular map of the *Hydra* nervous system using approximately 35,000 neural single cell transcriptomes, including differentiated neurons and cells undergoing neurogenesis. We surveyed for previously undetected diversity, and in addition to confirming 11 neuron subtypes (Siebert et al., 2019), we identified the first transcriptional evidence of neurons undergoing transdifferentiation in *Hydra*. We identified putative transcriptional regulators for all subtypes, performed ATAC-seq to identify the chromatin state of *Hydra* neurons, and built differentiation trajectories describing the transcriptional changes that underlie *Hydra* neurogenesis. This work represents the most detailed molecular description of the adult *Hydra* nervous system to date.

These results build on the work started in our previously published molecular map of the *Hydra* nervous system (Siebert et al., 2019). We increased the number of sequenced cells ten-fold and confidently identified 11 neuron subtypes. The number of identified subtypes has remained stable after increasing the number of profiled cells substantially, which suggests that our data reveals the entire complement of transcriptionally distinct neuron subtypes present in *Hydra*. Our neuron subtype analysis is also consistent with an scRNA-seq study from the Bosch lab in which approximately 1000 neurons were sequenced (Klimovich et al., 2020). A benefit of our larger data set is the higher capture rate of transition states (which are rare as compared to differentiated neurons), enabling a more complete reconstruction of the developmental trajectories that give rise to different neuronal subpopulations. For instance, we identified three transdifferentiation states: transdifferentiation clusters (td) 1, 2, and 3. In addition to identifying individual metagenes for each of these transdifferentiation populations (Figure 2.10), cells from each

transdifferentiation population were sequenced in all 16 libraries, suggesting that these are biologically valid cells rather than a library-specific batch effect. Our favored hypothesis is that these clusters represent ec1A cells in the body column that are undergoing transdifferentiation into either ec1B cells at the oral end or ec5 cells at the aboral end. This hypothesis is supported by our findings from the FDL analysis (Figure 2.5). The ability of neurons to transdifferentiate has been previously shown in *Hydra* that lack ISCs (Bode et al., 1988; Koizumi and Bode, 1991, 1986), but this is the first evidence that this phenomenon occurs as part of the normal homeostatic maintenance of the *Hydra* nervous system.

Our analysis identified marker genes of all neuron subtypes as well as different progenitor states (Figure 2.9), which will enable the creation of reporter lines to interrogate nervous system development and regeneration in future studies. The new, neuron-enriched ATAC-seq data generated in this study will facilitate this effort by enabling the accurate identification of regulatory regions. For instance, in this study, we created a new mNeonGreen reporter line using the regulatory region of *tba1c (G019559*), which is predicted to be expressed pan-neuronally (Figure 2.3 D-E). However, despite transmission through the germline, we found that *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ did not express mNeonGreen uniformly in all neurons, suggesting that we did not capture the full regulatory region of *tba1c* (for validation of line, see Figure 2.7). When making this line, we used peaks from the whole animal 105 ATAC-seq data (Siebert et al., 2019) as a guide for identifying the regulatory region and cloned 1901 bp upstream of the *tba1c* transcription start site (TSS). Our new neuron-enriched ATAC-seq data identifies an additional peak ~3,000 bp upstream of the TSS that was not captured in our cloning process (Figure 2.7 N). This missing element of the regulatory region could explain why we do not have uniform

transgene expression across the *Hydra* nervous system in our *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ line. This demonstrates the utility of our neuron-enriched ATAC-seq libraries, which provide novel information about the regulatory regions of neuronal genes and will help guide the design of neuronal promoter lines in the future. As we have identified specific gene markers for every neuron subtype, this will ultimately allow us to create transgenic reporter lines for each neuron subtype and individually analyze the regulatory regions of each of *Hydra's* 11 neural subtypes.

We have uncovered the molecular underpinnings of the entire *Hydra* nervous system, including 48 TFs that are expressed only in neurons or neuronal progenitors (Figures 2.2 and 2.11). This includes TFs that are common to all neurons as well as more selectively expressed TFs that are likely involved in the specification of individual neuron subtypes. These TFs are representative of many TF families, such as: C2H2 Zinc Finger family (10/48), bHLH superfamily (8/48), Winged Helix superfamily (4/48), Homeobox family (4/48), high mobility group box superfamily (3/48), SMAD family (3/48), bZIP family (2/48), Bed-type Zinc Finger family (2/48), and GATA-type Zinc Finger family (1/48). Many of these TFs are also conserved in bilaterian neurogenesis pathways, including: *myc3* (*G003730*) (Knoepfler et al., 2002)*, neurog1/2/3 (G008286)* (Blader et al., 1997; Lee, 1997)*, sox3 (G001357)* (Bylund et al., 2003; Rogers et al., 2013)*, ndf1 (G011383)* (Lee et al., 1995; Miyata et al., 1999)*, atoh8 (G021588)* (Jarman et al., 1993)*, and creb1 (G019837)* (Dworkin et al., 2007). This suggests their conserved role in regulating nervous system development in the last common ancestor of bilaterians and cnidarians.

We observed differences in TF expression patterns between endodermal and ectodermal neurons that may reflect different developmental strategies. The three endodermal neural subtypes (en1, en2, and en3) each selectively express one or two TFs (Figure 2.2), suggesting that these TFs act

as terminal selectors of the endodermal neural states; this is a similar strategy described for specification of *C. elegans* neurons (Doitsidou et al., 2018; Hobert, 2016; Patel and Hobert, 2017). By contrast, ectodermal neurons express a larger number of neuron-specific TFs and while each subtype expresses a unique set of these TFs, ec2 is the only neuron subtype to express unique TFs (*G018876* and *noto*) (Figure 2.2). This suggests that the ectodermal neurons are specified by combinatorial gene control and may also reflect their capacity to transdifferentiate between ectodermal neuronal subtypes.

To better understand the specification of *Hydra* neurons, we performed trajectory inference using URD and built a branching differentiation trajectory to describe the process by which all 11 neurons develop from the multipotent ISCs (Figure 2.4 A). Our data suggest that the first decision point of neuronal differentiation is the choice between endodermal and ectodermal fate. This may reflect the unique biology of these progenitors, given that the endodermal progenitor cells have to cross the ECM to populate the endodermal nerve net. Further, we note that groups of ectodermal neurons hypothesized to be part of the same neural circuits (RP1: ec3A, ec3B, ec3C and CB: ec1A, ec1B, ec5) (Dupre and Yuste, 2017; Keramidioti et al., 2023; Siebert et al., 2019) share progenitor states, with the exception of ec5.

All together, these data provide a comprehensive transcriptional description of the homeostatic *Hydra* nervous system. These data are a hypothesis generator and critical starting point for functional studies of nervous system development, regeneration, and function.

ACKNOWLEDGEMENTS

We would like to thank Rob Steele for providing the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ line, which was foundational to this study; the Juliano lab, Farrell lab, and David lab for invaluable discussions and feedback; Bruce Draper, Rob Steele, and Stefan Siebert for manuscript review and feedback; Clara Stuligross for analysis help, writing assistance, and moral support; Yohana Berhe for help with confocal images used in Figure 2.7; Diana Burkart-Waco, Siranoosh Ashtari, and Emily Kumimoto from from the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) for technical advice and assistance with library preparation and sequencing; Bridget McLaughlin and Jonathan Van Dyke from the UC Davis Flow Cytometry Shared Resource Laboratory (supported by NCI P30 CA093373 (Comprehensive Cancer Center), S10 OD018223 (Astrios Cell Sorter), S10 RR 026825 (Fortessa Cytometer), and NIH NCRR C06-RR12088 (Tupper Hall room 3425 sorting suite renovations)) for technical advice, assistance with FACS sorting, and good conversation. This work was supported by a National Institutes of Health (NIH) Grant R35 GM133689 (to CEJ), a W.M. Keck Foundation Grant (to CEJ) and a National Institutes of Health (NIH) Grant ZIAHD008997 (to JAF). ASP and HML were supported in part by an NIH T32 predoctoral training program in Molecular and Cellular Biology (GM007377).

AUTHOR CONTRIBUTIONS

ASP, JFC, JAF, and CEJ designed research; ASP, SM, and CND performed research; ASP, BDC, SM, and CND collected data; ASP, JFC, HML, SM, CND, and JAF analyzed data; ASP and CEJ wrote the initial manuscript draft, ASP, JFC, HML, BDC, CND, JAF, and CEJ reviewed and edited the manuscript; and CEJ, JAF acquired funding.

MATERIALS AND METHODS

Data Access

Data and all computational analyses conducted as part of this study are available at *https://doi.org/10.25338/B83S8C*

Generation of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic strain

Identification of pan-neuronal gene *tba1c* (*G019559*)

Genes expressed in all neurons were identified from the *Hydra* single cell RNA-seq dataset (data set ds.ds.genome (Siebert et al., 2019)) using the Seurat function *FindAllMarkers* (min.pct=0.50, other default parameters unchanged). Genes were filtered by avg-log fold change and the expression patterns of the top 10 genes were visualized using tSNE plots to identify genes with expression only in the neurons. The top three candidates were cloned into reporter constructs and tested for transgenic expression. Of these, only *tba1c* produced a transgenic animal with neuronal expression.

Cloning the regulatory region of *tba1c* (*G019559*)

To clone the *tba1c* regulatory region, we used gene models and peaks identified from published *Hydra vulgaris* strain 105 gene models and ATAC-seq data (Siebert et al., 2019). We cloned 1901 bp upstream of the *tba1c* transcription start site, capturing all identified peaks. The region was PCR amplified from *Hydra vulgaris* strain 105 gDNA using Phusion™ High-Fidelity DNA Polymerase (ThermoFisher Scientific, #F530S) with a 55°C annealing temperature and a 60°C extension temperature. Restriction sites for BamH1 and Xba1 were added onto the 5' ends of the F and R primers for ease of cloning upstream of mNeonGreen in the plasmid. After gel extraction, DNA was digested at 37°C for 10 minutes with BamH1 and Xba1 FastDigest

enzymes (ThermoFisher Scientific, #FD0054, #FD0684) and then purified using Zymogen DNA clean and Concentrator (#D4003). The purified DNA was ligated into the parent plasmid backbone (BsAmp_mNG_tdT_backbone) for four hours using Promega 10X fast ligase buffer (Promega, #M1801) at a concentration ratio of 2.5:1 insert:backbone. After transformation into competent DH5α bacteria, DNA was amplified via miniprep (Qiagen, #27106) and verified by Sanger sequencing. The reporter plasmid was amplified via maxiprep (Qiagen, #12162) and was eluted with autoclaved MQ water in preparation for microinjection into *Hydra* embryos.

Injecting the *Tg(tba1cmNeonGreen)*$^{cj1\text{-}gt}$ plasmid in *Hydra* embryos

Injections were performed as previously described (Juliano et al., 2014c) with the following modifications: 1) injection solution was prepared by mixing 1 µL 0.5% phenol red (Sigma P0290-100ML) with 6 uL plasmid DNA solution prior to centrifugation and 2) embryos were fertilized for 1-2 hours prior to injection. The plasmid injection solution was injected into *Hydra vulgaris* AEP 1-cell stage embryos using an Eppendorf FemtoJet 4x and Eppendorf InjectMan NI 2 microinjector (Eppendorf; Hamburg, Germany) under a Leica M165 C stereo microscope (Leica Microscopes, Inc; Buffalo Grove, Il). Two hatchlings with mosaic mNeonGreen neuronal expression were obtained and were propagated by asexual reproduction. To obtain a fully transgenic line via germline transmission, male and females from *Tg(tba1cmNeonGreen)*$^{cj1\text{-}in}$ were crossed to generate an F1 population.

Validation of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line

The hvCADab antibody has been shown to stain all nerve cells in *Hydra* (Keramidioti et al., 2023). Immunostaining with hvCADab was carried out as described previously (Keramidioti et

al., 2023). To improve antibody access to ectoderm and endoderm, fixed animals were cut in half or in smaller pieces. Following staining with primary and secondary antibodies, specimens were mounted and imaged with a Leica SP5 scanning confocal microscope using 20X and 63X objective lens. Image stacks were
processed with ImageJ (Abràmoff et al., 2004) to score pan-neuronal cells and hvCADab positive nerve cells.

Preparation and Collection of Cells using Fluorescence Activated Cell Sorting (FACS) for scRNA-seq and ATAC-seq

Dissociation of *Hydra* into single cells

Prior to dissociation, 40-45 starved, asexual polyps were washed 3x in sterile *Hydra* medium in 1.5 mL eppendorf tubes. *Hydra* medium was removed and ~75 units/mL Pronase E (VWR, E629-1G) in ~1 mL room temperature *Hydra* Dissociation Medium (DM) (5 mM CaCl2 2H20, 1 mM MgSO4 7H20, 2.8 mM KCl, 2.6 mM HEPES, 0.67 mM Na2HPO4, 0.44 mM KH2PO4, 5 mM Na Pyruvate, 5 mM Na3 Citrate 2H20) was added (Greber et al., 1992; Siebert et al., 2019). *Hydra* were dissociated for 90 minutes at room temperature (22-24°C) with gentle agitation on a nutator.

Following dissociation, cells were transferred to a small petri dish with 2 mL room temperature DM. Cells were gently pipetted up and down 5-10 times with a 1000 µL pipette to aid in final tissue separation and were strained through a pre-soaked 70 µM Corning® cell strainer (Sigma #CLS431751) into a 50 mL conical tube containing 1 mL DM. Tubes were tilted at an angle to ensure the cell solution gently slid down the side of the tube to prevent additional cell stress. The

petri dish and strainer were each rinsed with an additional 1 mL of DM. Cells were spun down for 5 min at 300 G in a centrifuge (Eppendorf Centrifuge 5702) with the brake turned off. After the first spin, the supernatant was carefully removed and cells were resuspended in 2 mL DM. If several tubes of the same *Hydra* strain were undergoing dissociation, they were combined at this step. Tubes were spun down for an additional 5 min at 300 G in a centrifuge with the break turned off. Then, supernatant was removed and the cell pellet was gently resuspended in 1 mL DM before a final filter through a pre-soaked 40 µM Corning® cell strainer (Sigma # CLS431750) into a 50 mL conical tube. Tubes were tilted at an angle to ensure the cell solution gently slid down the side of the tube to prevent additional cell stress

Collection of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ cells for 10x Single Cell RNA-seq

One scRNA-seq library was collected from asexual, bud-free *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}F1}$ *Hydra* (Table 2.1). *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (Keramidioti et al., 2023) was used as a positive control for FACS sorting and *Hydra vulgaris* AEP were used as a negative control. mNeonGreen positive cells were sorted using a 100 µM nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter, Miami, Fl, USA) into 400 µL of dissociation medium + 0.01% BSA. Cells were then spun down at 300 G for 5 minutes and resuspended in 50 µL of dissociation medium + 0.01% BSA. Cells were then processed for sequencing using the Chromium Next GEM Single Cell 3' kit v3.1 (10x Genomics) for sequencing according to manufacturer's instructions at the University of California, Davis Sequencing Core Facility. The chip was loaded with a targeted recovery of 10,000 cells and 11,183 cells were sequenced.

Collection of *Tg(actin1:GFP)*$^{rs3\text{-}in}$ cells for 10x Single Cell RNA-seq

Three scRNA-seq libraries were collected from asexual *Tg(actin1:GFP)*$^{rs3\text{-}in}$ animals (Keramidioti et al., 2023) (Table 2.1). One library was collected from bud-free animals and two libraries (*Tg(actin1:GFP)*$^{rs3\text{-}in}$ budding1 and *Tg(actin1:GFP)*$^{rs3\text{-}in}$ budding2) were collected from budding animals. Previously published gates (Siebert et al., 2019) were used for FACS sorting and *Hydra vulgaris* AEP were used as a negative control. GFP positive cells were sorted using a 100 µM nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter, Miami, Fl, USA) into 400 µL of dissociation medium + 0.01% BSA. Cells were then spun down at 300 G for 5 minutes and resuspended in 50 µL of dissociation medium + 0.01% BSA. Cells were then processed for sequencing using the Chromium Next GEM Single Cell 3' kit v3.1 (10x Genomics) for sequencing according to manufacturer's instructions at the University of California, Davis Sequencing Core Facility. The chip was loaded with a targeted recovery of 10,000 cells; the non-budding library contained 7,988 sequenced cells and the two budding libraries contained 7,374 and 6,124 cells.

Collection of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ cells for ATAC-seq

Two ATAC-seq libraries were collected from asexual, bud-free *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}F1}$ *Hydra* (Table 2.3). *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (Keramidioti et al., 2023) was used as a positive control for FACS sorting and *Hydra vulgaris* AEP were used as a negative control. mNeonGreen positive cells were sorted using a 100 µM nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter, Miami, Fl, USA) into 400 µL of dissociation medium + 0.01% BSA in DNA LoBind® Tubes (Eppendorf, #0030122348). Cell counts were validated using Hoechst staining and a

Fuchs-Rosenthal hemocytometer (*www.incyto.com*, DHC-F01-5). Prior to sorting, one additional mL of DM was added to the dissociated cells to aid in sorting (bringing the total to 2 mL).

Collection of *Tg(actin1:GFP)*$^{rs3\text{-}in}$ cells for ATAC-seq

Four ATAC-seq libraries were collected from asexual, non-budding *Tg(actin1:GFP)*$^{rs3\text{-}in}$ animals (Keramidioti et al., 2023) (Table 2.3). Previously published gates (Siebert et al., 2019) were used for FACS sorting and *Hydra vulgaris* AEP were used as a negative control. GFP positive cells were sorted using a 100 µM nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter, Miami, Fl, USA) into 400 µL of dissociation medium + 0.01% BSA in DNA LoBind® Tubes (Eppendorf, #0030122348). Cell counts were validated using Hoechst staining and a Fuchs-Rosenthal hemocytometer (*www.incyto.com*, DHC-F01-5). Prior to sorting, one additional mL of DM was added to the dissociated cells to aid in sorting (bringing the total to 2 mL).

10x Genomics Single Cell RNA Sequencing (scRNA-seq)

scRNA-seq Read Mapping

scRNA-seq libraries collected via 10x Single Cell Genomics were processed using cellranger-4.0.0 tools according to the manufacturer's guidelines and aligned to the *Hydra vulgaris* AEP transcriptome (Cazet et al., 2023). An average overall transcriptome alignment rate of 59.9% was obtained (*Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$, 62.5%, *Tg(actin1:GFP)*$^{rs3\text{-}in}$ non-budding 58.6%, *Tg(actin1:GFP)*$^{rs3\text{-}in}$ budding1 60.4%, *Tg(actin1:GFP)*$^{rs3\text{-}in}$ budding2 58.1%). For the scRNA-seq data collected from Siebert et al. and realigned to the *Hydra vulgaris* AEP genome (Cazet et al., 2023), we generated a table containing raw reads from interstitial stem cells, neuronal progenitors, and differentiated neurons based on cluster identity. Cluster analysis was performed

using the Seurat v 4.0.5 package (Butler et al., 2018; Hao et al., 2021a; Satija et al., 2015; Stuart et al., 2019).

Quality Control

Individual Seurat objects were generated from each library and were filtered as previously described to retain cells with between 300 - 7,000 uniquely expressed genes, 500 - 50,000 transcripts, and less than 5% mitochondrial reads (Satija et al., 2015; Siebert et al., 2019). Raw and filtered data were visualized using violin plots (Figure 2.8). Because the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ line contains a small population of transgenic nematocytes and gland cells in addition to neurons, we performed a basic cluster analysis of the three libraries collected using *Tg(actin1:GFP)*$^{rs3\text{-}in}$ animals. After clustering, expression plots of nematocyte and gland cell molecular markers (identified in (Siebert et al., 2019)) were generated and any clusters expressing those markers were removed. Next, we generated module scores for each object based on the list of genes identified in a preliminary analysis as “stress markers”. These scores were used as regression variables during object integration. Note: We did not perform doublet removal using mutually exclusive NMF modules as described in (Siebert et al., 2019) to avoid removing cells that were undergoing transdifferentiation, which could potentially appear to be doublets to most computation doublet-removal approaches.

Initial Library Integration, Principal Component Analysis (PCA) ,and Cluster Analysis

We integrated all of our datasets (12 previously published and 4 newly generated) using the SCTransform pipeline (Hafemeister and Satija, 2019). We excluded the previously published Drop-seq library D06_FM_S1 because it contained too few neuronal cells to properly integrate

(68 cells). First, SCTransform was used to normalize the data sets, regressing against the stress marker module scores and percent of reads that mapped to mitochondrial RNA. Next, data were integrated using commands Seurat::*SelectIntegrationFeatures* (*nfeatures* = 3000), Seurat::*PrepSCTIntegration*, Seurat::*FindIntegrationAnchors*, and Seurat::*IntegrateData.* We then performed principal component analysis (PCA) using the command Seurat::*RunPCA*. To determine the number of principal components to use for downstream analysis, we identified the point where the percent change in variation between consecutive PCs is less than 0.1%, as suggested by (*https://hbctraining.github.io/scRNA-seq/lessons/elbow_plot_metric.html*). We then generated a preliminary clustering and UMAP using the Louvain approach (Seurat::*FindClusters*, (*dims.use* = 1:25), Seurat::*FindNeighbors* (*resolution* = 0.8), Seurat::*RunUMAP*, (*dims.use* = 1:25)).

We then eliminated any non-neuronal or high-stress clusters, based on their expression of module scores generated from stress markers, mitochondrial RNAs, and previously identified molecular markers for neurons, progenitors, interstitial stem cells, nematocytes, and gland cells (Siebert et al., 2019). Only one cluster expressed stress markers and was subsequently removed.

Second Library Integration, Principal Component Analysis (PCA), and Cluster Analysis

We then repeated similar processing steps after eliminating non-neuronal and high-stress clusters: normalizing with Seurat::*SCTransform*, regressing against the same values, integrating with Seurat::*SelectIntegrationFeatures* (*nfeatures* = 3000), Seurat::*PrepSCTIntegration*, Seurat::*FindIntegrationAnchors*, and Seurat::*IntegrateData,* and performing principal component analysis with the command Seurat::*RunPCA*. We then reclustered the data using 21 PCs and

resolution 0.8, and re-generated a UMAP projection using 21 PCs, with parameters *min.dist* = 0.25 and *spread* = 0.7.

To validate the clustering, we plotted molecular markers for each neuron subtype, neuronal progenitors, and interstitial stem cells. All but four clusters were identified by this method. Markers for the four clusters with unknown identities were plotted in the whole animal single cell data set (Cazet et al., 2023; Siebert et al., 2019) to determine a possible origin of the cells. Only one unknown cluster's markers were stress related and had no expression in the whole animal data set, and thus was subsetted out for removal. The remaining three unknown clusters, td1, td2, and td3, all expressed markers found in ec1A, ec1B, and ec5 neurons, and hence were not removed from the data set. In order to visualize expression information downstream, Seurat RNA assay was log normalized using the command Seurat::*NormalizeData* with a scale factor of 10,000 and used as the basis for dot plots and other gene expression plots.

scRNA-seq Data Analysis: Identification of Putative Transcription Factors Expressed in Neurons

To identify transcription factors expressed specifically in neurons, we visualized the expression patterns of 812 genes with DNA binding domains that have detectable expression in the adult *Hydra* polyp (Cazet et al., 2023; Siebert et al., 2019). Gene expression was screened based on apparent neuronal specificity in gene expression plots made using Seurat*::DotPlot.* Neuronal expression was confirmed by comparison to whole-animal single-cell data from (Cazet et al., 2023; Siebert et al., 2019). This resulted in a list of 48 putative transcription factors.

ATAC-seq

Library Preparation

To generate neuron enriched ATAC-seq libraries, we used a modified version of the OMNI_ATAC protocol (Cazet et al., 2023; Corces et al., 2017; Siebert et al., 2019). Briefly, dissociated cells were collected via FACS (as described above) and spun down in an Eppendorf 5424R centrifuge at 1000 G for 5 minutes at 4°C. Cell pellets were resuspended in 50 µL of freshly made, chilled resuspension buffer (RSB) (10 mM Tris-HCl - pH 7.4, 10 mM NaCl, 3 mM MgCl2) containing 0.1% tween, 0.1% NP-40, and 0.01% digitonin. After three minutes of incubation on ice, lysis was stopped by adding 1 mL of chilled RSB plus 0.1% tween and tubes were mixed via inversion 3x. Nuclei were spun down at 500 G for 10 min at 4°C. After removing the supernatant, nuclei were resuspended in freshly made tagmentation mix [1x TD buffer (Illumina, San Diego, CA #20034197), 33% PBS, 0.01% digitonin, 0.1% tween-20, 5 ul TDE1 (Illumina #20034197)] and incubated for ~30 min shaking at 1,000 RPM at 37°C on an Eppendorf Thermoshaker C. Tagmentation was ended by adding 250 µL of PB buffer from the Qiagen MinElute PCR Purification Kit (Qiagen, #28004) and samples were stored at -20°C for up to 2 weeks.

Tagmented DNA was brought to room temperature and purified using the Qiagen MinElute PCR Purification Kit (Qiagen, #28004) using the manufacturer's standard protocol, with a final elution in 21 µL of EB buffer. Libraries were then amplified with 2X NEBNext master mix (NEB, Ipswitch, MA M0541S) using cycle numbers determined by qPCR as described in the standard ATAC-seq protocol (Table 2.3, Column "Total PCR Cycles for Amplification") (Buenrostro et al., 2013).

Agencourt AMPure XP beads (Beckman Coulter, Pasadena, CA #A63881) were then used to purify libraries and restrict fragment sizes to between 100 and 700 bps. DNA concentration was determined using QuBit dsDNA HS assay (ThermoFisher Scientific, #Q32851) and library size was determined using the Bioanalyzer High-Sensitivity DNA kit at the UC Davis DNA Core (Agilent Cat # 5067-4626). Of the 5 biological replicates collected from *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ *Hydra*, two replicates were chosen to sequence based on QuBit concentration and Bioanalyzer traces. Of the 6 biological replicates collected from *Tg(actin1:GFP)*$^{rs3\text{-}in}$ *Hydra*, four replicates were chosen to sequence based on QuBit concentration and Bioanalyzer traces. Libraries were then pooled at roughly equimolar proportions and sequenced on an Illumina HiSeq4000 (*Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$) or Illumina HiSeq X Ten (*Tg(actin1:GFP)*$^{rs3\text{-}in}$) using 2x150bp reads.

Genome Alignment and Peak Calling

Raw sequencing data were filtered using Trimmomatic to remove sequencing adaptors, as well as low-quality and unpaired reads (Bolger et al., 2014). Filtered reads were then mapped to the *Hydra vulgaris* AEP genome (Cazet et al., 2023) using Bowtie2 (Langmead and Salzberg, 2012), followed by filtering of improperly or ambiguously mapped reads using SAMtools (Li et al., 2009). PCR duplicates were then identified using Picard Tools (broadinstitute.github.io/picard) and removed with SAMtools.

Peak Calling was performed using code adapted from (Cazet et al., 2023) which was originally modified from the ENCODE consortium's ATAC-seq analysis pipeline (encodeproject.org/atac-

seq) (Landt et al., 2012). First, reads from the completed mapping pipeline were indexed, sorted, and centered over the transposase binding site by shifting plus strand reads +4 bp and minus strand reads −5 bp using deepTools (Ramírez et al., 2016). Replicates were then 1) divided into two self-pseudoreplicates and 2) pooled and split into biological psuedoreplicates. Psuedoreplicates were used to generate consensus lists of biologically reproducible peaks by using the irreproducible discovery rate (IDR) framework (Li et al., 2011) to identify peaks that were reproducible (IDR score ≤0.1) across at least two pairwise comparisons of biological replicates in each transgenic line. Peaks were then called using Macs2 (Zhang et al., 2008) with a permissive p-value cutoff of 0.1.

For downstream analyses, bigwig files, transcription start site (TSS) enrichment scores, self-consistency, and rescue ratios were calculated (as performed in (Cazet et al., 2023; Siebert et al., 2019)). We compared the six neuron-enriched libraries collected in this study with three previously published whole *Hydra vulgaris* strain AEP ATAC-seq libraries (Cazet et al., 2023) and found that our sorted ATAC-seq libraries were of equivalent high quality as compared to our published whole animal libraries (Table 2.5).

ATAC-seq Data Analysis: Identifying Differentially Expressed Peaks

Data were then analyzed using DiffBind to calculate normalized read counts across all peaks in the *Tg(actin1:GFP)*$^{rs3\text{-}in}$, *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$, and AEP samples. Read counts were saved as bed files. UROPA was then used to annotate all ATAC-seq peaks based on nearest TSS (Kondili et al., 2017). Next, we used edgeR to identify differentially accessible peaks based on the read counts generated by DiffBind between the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$, *Tg(actin1:GFP)*$^{rs3\text{-}in}$, and AEP data sets. The results of this script are saved as an RData object

and individual results tables. Plots visualizing ATAC-seq data were generated using Gviz (Hahne and Ivanek, 2016).

ATAC-seq Data Analysis: Identifying Peaks Enriched Near Neuronal Genes

Genes positively (NeuroG) or negatively (nonNeuroG) associated with differentiated neurons were identified from the whole animal single cell data (Cazet et al., 2023) using Seurat. Peaks associated with NeuroG or nonNeuroG genes were identified from the pairwise comparison results *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$/AEP and *Tg(actin1:GFP)*$^{rs3\text{-}in}$/AEP of the ATAC EdgeR analysis and the mean log fold change was calculated for each. Results of log fold change for each NeuroG or nonNeuroG associated peak and the mean log fold change were visualized in a violin plot generated by ggplot2. A two-sided t-test was performed to determine the significance of differentially accessible peaks between NeuroG and nonNeuroG genes

Trajectory Analysis

To explore both the primary differentiation pathway of neurons from ISCs as well as potential transdifferentiation events from one differentiated neuron type to another, we took two complementary approaches for building developmental trajectories: URD (Farrell et al., 2018) to describe primary differentiation from ISCs and structure-agnostic Force Directed Layouts (FDL) (Fruchterman and Reingold, 1991) to describe potential transdifferentiation events.

URD Analysis of Primary Differentiation Events

URD (Farrell et al., 2018) was used to generate trajectories describing the differentiation of *Hydra's* 11 neuron subtypes from ISCs. In an initial analysis, we observed strong batch effects

between the data generated using 10x scRNA-seq and Drop-seq that inhibited the proper formation of connections between these cells during the calculation of cell-to-cell transitions. Because the Drop-seq data only comprised a minority of the dataset (5,400 cells), we excluded the Drop-seq data from our trajectory analysis and used only the cells generated using 10x scRNA-seq. Additionally, since our goal was to focus on primary differentiation, and since URD does not reconstruct cyclical structures, we also removed the putative transdifferentiation events from our data.

We then used the SeuratToURDv3 function to generate an URD object, and manually added in the integrated data from our Seurat object (*urd.obj@logupx.data* <- *seurat.obj@assays$integrated@data*) Next, we used the URD::*urdSubset* function to eliminate the Drop-seq data and putative transdifferentiation events as described above.

Removal of Outlier Cells and Doublets

Outlier cells are poorly connected to the main data set and often disrupt trajectory reconstruction. We identified 209 outlier cells (based on their distance to their nearest neighbors) using the function URD::*knnOutliers*, with parameters *x.max* = 86, *slope.r* = 0.12, int.r = 79, *slope.b* = 1.1, and *int.b* = 7.75.

Doublets are formed at low frequencies when using a droplet-based scRNA-seq technique when multiple cells are encapsulated in a single droplet. Since these can create spurious connections that interfere with trajectory inference, we aimed to remove them by identifying cells with co-expression of typically mutually exclusive gene expression programs (Siebert et al., 2019). We

normalized NMF scores across all cells in the object by scaling such that each module's expression ranged from 0-1 across all cells. Modules were assessed for cell-type specificity by visually confirming their expression on the UMAP projection (Figure 2.10). Then, pairs of modules that should be mutually exclusive and the cells that co-express them were identified using the command URD::*NMFDoubletsDetermineCells*, with parameters *module.expressed.thresh* = 0.2, *frac.overlap.max* = 0.07, *frac.overlap.diff.max* = 0.15. A total of 1,257 cells were identified that expressed pairs of modules at a high level. Outliers and doublets were removed from the URD object using command URD::*urdSubset.*

Construction of Branching Trajectory

To calculate pseudotime and identify branching trajectories leading to our 11 neuronal cell types, we calculated a diffusion map using the command URD::*CalcDM* (which draws on the R package *destiny*) (Angerer et al., 2016), using the parameters *knn*=100, *sigma.use* = local, *distance* = cosine. These parameters were chosen by comparing results from several parameters. When selecting the diffusion map, we looked for (1) strong connections between differentiated neurons and progenitors, and (2) low promiscuity between different groups of terminally differentiated cells that we hypothesized shouldn't be related (and for which we did not see cells with intermediate gene expression states).

We next calculated pseudotime using the interstitial stem cells, as determined by expression of *Hydra* ISC marker (*G002332*) (Siebert et al., 2019), as the "root", or starting point of the tree. Terminal neural populations were chosen from Infomap-Jaccard community detection clustering produced by URD::*graphClustering*, using parameters *do.jaccard* = T, *method* = "Infomap",

num.nn = 120. Tip clusters were selected based on (1) their late pseudotime as assigned by URD and (2) several differentially expressed genes. Pseudotime was then computed using the commands URD::*floodPseudotime* with the following parameters: *n* = 50, *minimum.cells.flooded* = 2. Logistical parameters for biasing the transition matrix were determined using the following parameters: *optimal.cells.forward* = 0, *max.cells.back* = 200. Since *Hydra* are constantly replenishing their tissues in a homeostatic manner, cell density along developmental processes varies widely, with a large number of transcriptionally similar differentiated cells and much smaller numbers of cells in transition. Thus, we used a larger *max.cells.back* value in an attempt to force the random walks to visit, and not bypass, cells of similar pseudotime when trajectory reconstruction was proceeding through regions of high cell density.

Biased random walks were then performed to determine the cells visited from each terminal population in the data using the following parameters: *n.per.tip* = 50000*, root.visits* = 1. Cells visited by random walks were visualized using a UMAP projection to ensure that the majority of the data was visited and that the tips chosen were well connected to the data and followed a specific path through it. In cases where multiple tips were tested for a subtype, chosen tips were combined prior to running the tree using URD::*combineTipVisitation*. The branching tree was then constructed using URD::*buildTree* with the following parameters: *divergence.method* = "preference"*, save.all.breakpoint.info* = TRUE*, cells.per.pseudotime.bin* = 25*, bins.per.pseudotime.window* = 8*, p.thresh* = 1e-6*, and min.cells.per.segment* = 10.

Determining Temporal Ordering of Genes During Differentiation

To visualize the temporal sequence of genes expressed during the specification and differentiation of individual neuronal subtypes, we used spline curves. Genes were considered that were expressed (*i.e.* > 0) in at least 1% of the population. Splines were calculated using tree segments containing cells belonging to the differentiation pathway of interest (e.g. all segments in ec3A differentiation were included). A spline curve was fitted to the mean expression vs. pseudotime relationship of each gene using the URD::*geneSmoothFit* function with the parameter *spar* = 0.875. Sets of genes were plotted using URD::*plotSmoothFit* to determine temporal expression; genes with overlapping expression domains were chosen to validate transition states predicted by the URD trajectory.

Force Directed Layout (FDL) Analysis of Transdifferentiation Events

Since URD assumes an underlying tree-like topology when determining developmental trajectories, and the putative transdifferentiation events we observed would violate that structure, we used force-directed layouts (FDLs) as an alternative method that is structure-agnostic to visualize transdifferentiation events (Farrell et al., 2018; Fruchterman and Reingold, 1991). To do so, we first isolated groups of neurons that we hypothesized to undergo transdifferentiation with each other and processed each group individually. In addition to performing FDL on groups of cells we hypothesized to undergo transdifferentiation, we also ran a control using a group of neuron subtypes with little transcriptional similarity. We used the command URD::*calcEnforcedKNN* to calculate weighted nearest neighbor networks, using distance in PCA space. When investigating the connections between ec1A, ec1B, ec5, and td1-3, we considered 25 neighbors by running URD::*calcEnforcedKNN* using parameters: *embedding* = "pca",

dims.use = 1:40, *nn.start* = 25, *nn.final* = NULL, *local.multiplier* = NULL, *mutual.only* = F. When investigating the connections between ec3A, ec3B, and ec3C, we considered neighbors up to 5x as distance as each cell's nearest neighbor (maximum 500 neighbors) by running URD::*calcEnforcedKNN* using parameters: *embedding* = "pca", *dims.use* = 1:40, *nn.start* = 500, *nn.final* = NULL, *local.multiplier* = 5, *mutual.only* = F. The calculated adjacency matrices were saved as Gephi spreadsheets using the R *igraph* package, and force directed layouts were calculated using the Gephi implementation of ForceAtlas2 (Jacomy et al., 2014), with parameters *--targetChangePerNode 0.5 --targetSteps 10000 --2d --format txt --seed 9481 --barnesHutTheta 1.2 --barnesHutUpdateIter 1 --jitterTolerance 1 --scalingRatio 2 --gravity 1*.

Tree Validation using Double Fluorescent In Situ Hybridization

To validate several transition states that our trajectory predicted, we used double fluorescent in situ hybridization (FISH) (Siebert et al., 2019), focusing on pairs of genes with overlapping expression domains during three predicted transition states of ec3A differentiation.

Probe Generation

To generate labeled RNA probes for double fluorescence in situ hybridization, we cloned and sequenced PCR products for *Hydra* genes *bhlha15* (*G021353*), *gata3* (*G022640*), *hym355* (*G004115*), and ec3A marker *G021930*. Amplicons were generated using oligo-dT primed cDNA generated from *Hydra vulgaris* strain AEP. The reverse primer sequences included either T7 or SP6 promoter sequences (Table 2.6), allowing us to use purified PCR products as templates for in vitro transcription reactions. PCR products isolated using gel extraction (Qiagen #28506) and blunt end ligated using the Invitrogen Zero Blunt PCR Cloning Kit (#K2700-20).

DNA from successfully transformed colonies was extracted and sequenced and plasmids containing the correct insert were amplified using a Qiagen MiniPrep Kit (Qiagen, #27106). The plasmid DNA was then used as a template for a second round of PCR using the original primers (Table 2.6) to generate the template for in vitro transcription. Amplicons were gel extracted and 250 ng of gel purified DNA was used as a template in an in vitro transcription reaction using the Roche DIG RNA labeling kit (Sigma #11175025910). To allow for double labeling, probes were transcribed using either DIG-U-11 or FITC-U-12 (Sigma #11685619910) NTP labeling mixes (Table 2.6). Ambion RNAse-in was used as the RNAse inhibitor. Probes were then purified using the Zymogen RNA Clean and Concentrator-25 kit (Zymo Research Cat # R1017), diluted to 750 ng aliquots in 7.5 uL H20, and stored at -80°C until use.

Day 1: Fixing and Clearing Samples

For each double in situ, we used ~25 *Hydra vulgaris* AEP polyps that had been starved for two days, transferred to 1.5 mL eppendorf tubes, and washed three times in fresh *Hydra* Medium (HM). *Hydra* were relaxed at room temp (RT) for 1.5 minutes in 1 mL HM containing 2% urethane, and then fixed in 1 mL HM containing fresh, ice-cold 4% paraformaldehyde (PFA) for 1 hour gently rocking at RT. All subsequent steps were performed at RT in 1 mL of solution while gently rocking unless otherwise noted. Following fixation, PFA was removed with three 10 minute PBT (0.1% tween-20 in phosphate buffered saline, pH 7.4) washes. The tissue was then bleached by transferring the samples gradually to 100% MeOH using 5 minute washes in 33% MeOH/PBT followed by 66% MeOH/H20. The samples were incubated in 100% MeOH for 1 hr, with the MeOH refreshed at the 30 minute mark. To maximize bleaching, samples were then incubated overnight in fresh MeOH at -20°C.

Day 2: Preparing Samples for Hybridization and Adding Probes

Samples were rehydrated using 5 minute washes of 66% MeOH/H20 and 33% MeOH/PBT, followed by three 5 minute PBT washes. Tissue was then permeabilized in 10 ug/mL proteinase K in PBT for 5 minutes each. (Note: Tissue becomes extremely fragile when treated with proteinase K, so do not exceed permeabilization time or pipette too vigorously). Proteinase activity was quenched using a quick wash with glycine working solution (4 mg/mL glycine in PBT), followed by a 10 minute wash in fresh glycine working solution. Tissue was then washed three times for 5 minutes each in PBT. The samples were then washed twice in 0.1 M triethanolamine in PBT, once in 0.1 M triethanolamine in PBT containing 3 µl/ml acetic anhydride, once in 0.1 M triethanolamine in PBT containing 6 µl/ml acetic anhydride, then three times in PBT, all for 5 minutes each. (Note: Add acetic anhydride immediately before use). Next, tissue was refixed in 4% PFA in PBT for 1 hour. PFA was removed with three 5 minute PBT washes followed by two 5 minute washes in 2X SSC (300 mM NaCl and 30 mM sodium citrate).

In preparation for probe hybridization, samples were incubated in 50% 2X SSC/50% hybridization solution [HS; 50% formamide, 5x SSC (750 mM NaCl and 75 mM sodium citrate), 1x Denhardt's solution, 100 µg/mL heparin, 0.1% Tween-20, and 0.1% Chaps] for 10 minutes, starting first at RT and then transitioning to hybridization temperature at 56°C. All subsequent steps were carried out at 56°C using reagents that were preincubated at this temperature. The tissue was then incubated in HS for 10 minutes and then in HS containing 10 µL/mL sheared salmon sperm for 2 hours. To prepare the DIG- and FITC-labeled probes for hybridization, we added ~750 ng of each probe from the appropriate pairing (e.g. *bhlha15* + *gata3*) to 35 µL 50%

2X SSC/50% HS for a final volume of 50 µL. Probes were then denatured via incubation at 85°C for 5-10 minutes. The probes were then transferred to the hybridization oven, and after cooling to hybridization temperature, were added to 400 uL of HS for a total volume of 450 µL. All but ~50 µL of liquid was removed from each eppendorf tube containing the animals and 450 µL of HS + Probe was added to the appropriately labeled tube. Each tube was wrapped in parafilm to prevent evaporation and samples were then left to hybridize for ~65 hours with no agitation. To ensure even probe distribution, tubes were gently mixed once every 24 hours during the hybridization period.

Day 3: Probe Removal and anti-Fluorescein Primary Antibody Incubation

Probe was removed using a sequence of single, 5 minute washes in HS, 75% HS/25% 2X SSC, 50% HS/50% 2X SSC, and 25% HS/75% 2X SSC at 56°C. Samples were then washed twice with 2X SSC containing 0.1% CHAPS for 30 minutes each, with the first wash occurring at 56°C and the second at RT. Samples were then washed four times with MABT (100 mM maleic acid, 150 mM NaCl, 0.1% Tween 20, pH 7.5) for 10 minutes each. Tissue was then washed with MABT + 1% BSA for 1 hour at RT, followed by a 2 hour incubation in 500 uL of blocking solution (80% MABT + 1% BSA/20% sheep serum) at 4°C. Samples were then resuspended in a 1:2000 dilution of Anti-Fluo-POD (Sigma, #11426346910) and incubated overnight at 4°C without agitation.

Day 4: Detection of Fluorescein-Labeled Probe and anti-DIG Primary Antibody Incubation

Following primary antibody binding, samples were transitioned back to RT and excess antibodies were removed with two 20 minute MABT-BSA washes followed by 5 20 minute MABT washes. Samples were then washed twice for 5 minutes in 100 mM Borate Buffer (1:1 of

200 mM borate stock (pH 8.5): boric acid 200 mM, sodium chloride 75 mM, sodium tetraborate (borax) 25 mM) + 0.1% Tween. Samples were then stained using 75 uL of tyramide solution (100 mM Borate Buffer, 2% dextran sulfate, 0.1% Tween-20, 0.003% H2O2, 0.15 mg/mL 4-iodophenol in DMSO, 1:100 dilution of Alexa Fluor 488 (Invitrogen #B40953) or 594 tyramide reagent (Invitrogen #B40947) for 25 minutes. The reaction was stopped using four quick PBT washes and samples were incubated in 100 mM glycine (pH 2.0) for 10 minutes at RT. Samples were then washed five times for 5 minutes with PBT. Samples were then incubated for 2 hours in 500 uL of blocking solution (80% MABT + 1% BSA/20% sheep serum) at 4°C. Samples were then resuspended in a 1:2000 dilution of Anti-DIG-POD (Sigma, #11207733910) and incubated overnight at 4°C without agitation. (Note: After performing the tyramide reaction, samples should be kept in the dark for the remainder of the protocol).

Day 5: Detection of DIG-Labeled Probe and tissue mounting

Following primary antibody binding, samples were transitioned back to RT and excess antibodies were removed with two 20 minute MABT-BSA washes followed by 5 20 minute MABT washes. Samples were then washed twice for 5 minutes in 100 mM Borate Buffer (1:1 of 200 mM borate stock (pH 8.5): boric acid 200 mM, sodium chloride 75 mM, sodium tetraborate (borax) 25 mM) + 0.1% Tween. Samples were then stained using 75 uL of tyramide solution (100 mM Borate Buffer, 2% dextran sulfate, 0.1% Tween-20, 0.003% H2O2, 0.15 mg/mL 4-iodophenol in DMSO, 1:100 dilution of Alexa Fluor 488 or 594 tyramide reagent) for 25 minutes. The reaction was stopped using four quick PBT washes and samples were incubated in 100 mM glycine (pH 2.0) for 10 minutes at RT. Samples were then washed five times for 5 minutes with PBT. In preparation for imaging, samples were stained in 1:1000 Hoechst in PBT

for 30 minutes. (Note: can go more dilute with Hoechst staining). Samples were then dehydrated through a gradient of 30%, 50%, and 80% glycerol in PBT, each lasting at least an hour. (Note: A good measure is to wait for the animals to float in the glycerol before moving to the next solution). Animals were then mounted in 80% glycerol with 40 mM NaHC03 and kept at 4°C until ready to image.

Confocal Imaging of Fluorescent In Situ Hybridization Samples

Samples were imaged on a Zeiss LSM 980 with Airyscan 2 microscope housed at the Light Microscopy Core at UC Davis. All images were collected as Z stacks using the Zen Blue software with 20x air lens and the confocal setting with scan speed set to 8 and averaging set to 2. Pinhole size was set to 27.42 microns. Image size for all images is 0.207 microns per pixel in the X and Y directions and 0.440 microns in the Z direction. Two imaging tracks were used, one for collecting the Hoechst nuclear DNA stain and one for collecting the two (green and red) fluorescent in situ signals. Track 1 used the Hoechst 33258 setting with excitation/emission wavelength maxima of 352 and 455 nm, respectively. Track 2 used the Alexa Fluor 488 setting with excitation/emission wavelength maxima of 493 and 517 nm and the Alexa Fluor 594 setting with excitation/emission wavelength maxima of 590 and 618 nm.

Identification of gene co-expression modules using non-negative matrix factorization (NMF)

To identify sets of co-expressed genes (i.e., metagenes) in our neuronal scRNA-seq data, we performed non-negative matrix factorization (NMF) as implemented by the cNMF package (Kotliar et al., 2019). As input for this analysis, we used raw read counts for all neuronal and

interstitial stem cell transcriptomes from all 10X and Drop-seq libraries. The data were first processed by the 'prepare' function within the cNMF pipeline to normalize read counts and to exclude genes with low variability across cells. Next, because the optimal number of gene co-expression modules (specified by the parameter *k*) for a given dataset needs to be empirically determined, we used the 'factorize' function to perform independent NMF analyses for *k* values ranging from 5 to 100 by steps of 5. In addition, 200 independent analyses were performed for each *k* value to enable the evaluation of metagene reproducibility. The results for all *k* values were then compiled using the 'combine' function and robustness and accuracy of the results were evaluated using the 'k_selection_plot' function. We determined that the optimal value for k that minimized error while maximizing stability fell between the values of 25 and 35. Because our initial analysis was performed using steps of 5, we repeated our analysis to systematically evaluate all *k* values between 25 to 35. Based on this more granular analysis, we selected a *k* value of 27. We then used the 'consensus' function to generate a set of 27 consensus metagenes from the 200 independent runs performed for that *k* value, filtering out those results that were most dissimilar from the results found in the majority of independent analyses.

FIGURES

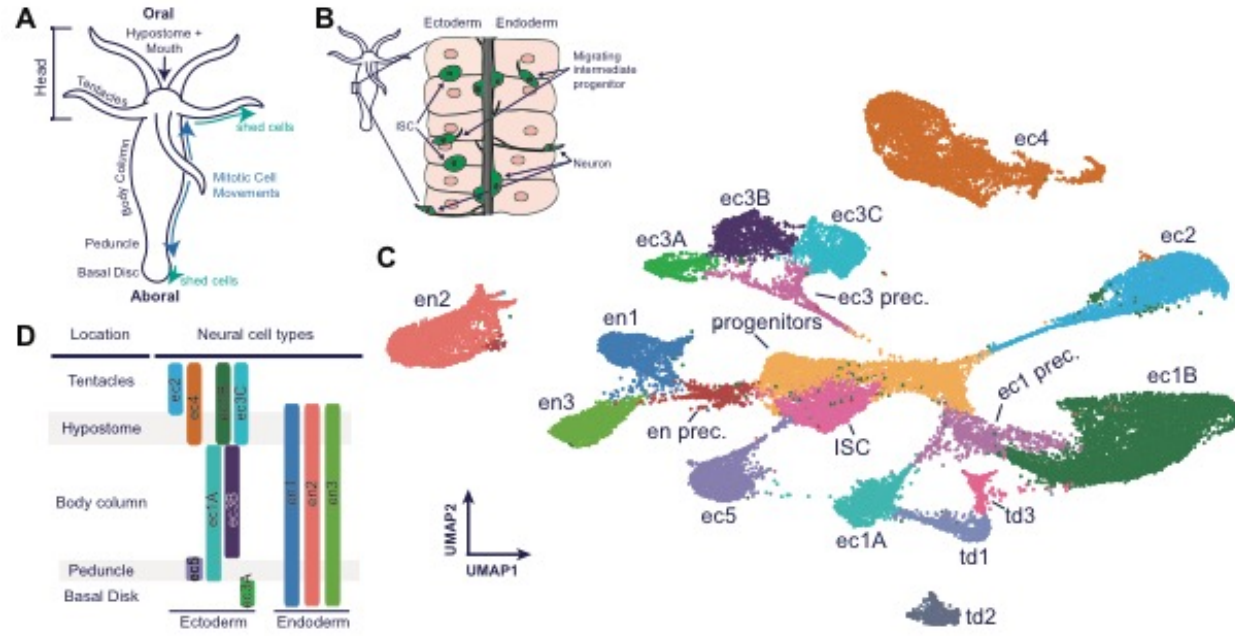


Figure 2.1. The *Hydra vulgaris* nervous system is composed of eleven transcriptionally distinct neuron subtypes. (A) The *Hydra* body is a radially symmetric hollow tube arranged around an oral-aboral axis. The hypostome, mouth, and tentacles (the "head") are located at the oral end and the peduncle and basal disk are located at the aboral end. These are connected by a body column consisting of dividing epithelial cells that generate passive tissue displacement towards the extremities, where the cells are eventually sloughed off (direction of tissue movement denoted with arrows). (B) *Hydra* consists of two epithelial cell layers, the endoderm and ectoderm, separated by an extracellular matrix (ECM). Neurons reside in the interstitial spaces of the epithelial cells to form two separate nerve nets, one in the ectodermal layer and one in the endodermal layer (Keramidioti et al., 2023). While the interstitial stem cells that give rise to neurons are found only in the ectoderm, intermediate neural progenitors migrate through the ECM to give rise to endodermal neuron subtypes. (C) 35,071 single cell transcriptomes were sequenced using a combination of Chromium Single-Cell Gene Expression (10x Genomics) (29,671 from this study) and Drop-Seq (5,400 from (Siebert et al., 2019)). UMAP representation of clustered cells annotated with cell state. ISC, interstitial stem cell; prec., precursors; ec, ectodermal; en, endodermal; td, transdifferentiation. The transdifferentiation clusters are likely intermediate states of transdifferentiation (see Figure 2.5 for details) (D) Spatial location of the 11 neuron subtypes along the *Hydra* body. Colors match the clusters in Figure 2.1 C. Figure adapted from (Badhiwala et al., 2021) except that the ec1A spatial location has been adjusted to include the peduncle based on evidence provided by (Noro et al., 2019).

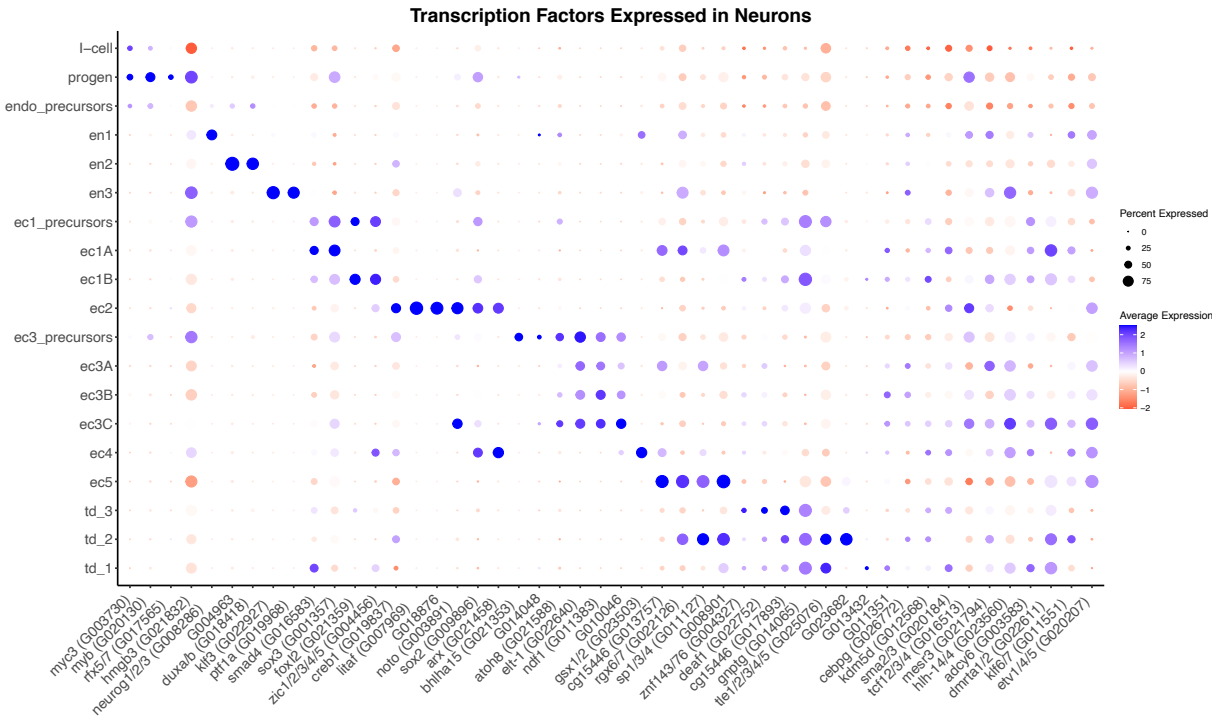


Figure 2.2. Forty-eight transcription factors (TFs) are specifically expressed in *Hydra* neurons. Dotplot representation of putative TFs uniquely expressed in the *Hydra* nervous system, including neuronal progenitor cells and each of *Hydra's* 11 neuron subtypes. Gene expression values are depicted by average expression across all cells (dot color) and percent expression within each cluster (dot size). Putative TFs were identified using PFAM annotations (data from (Cazet et al., 2023; Siebert et al., 2019)).

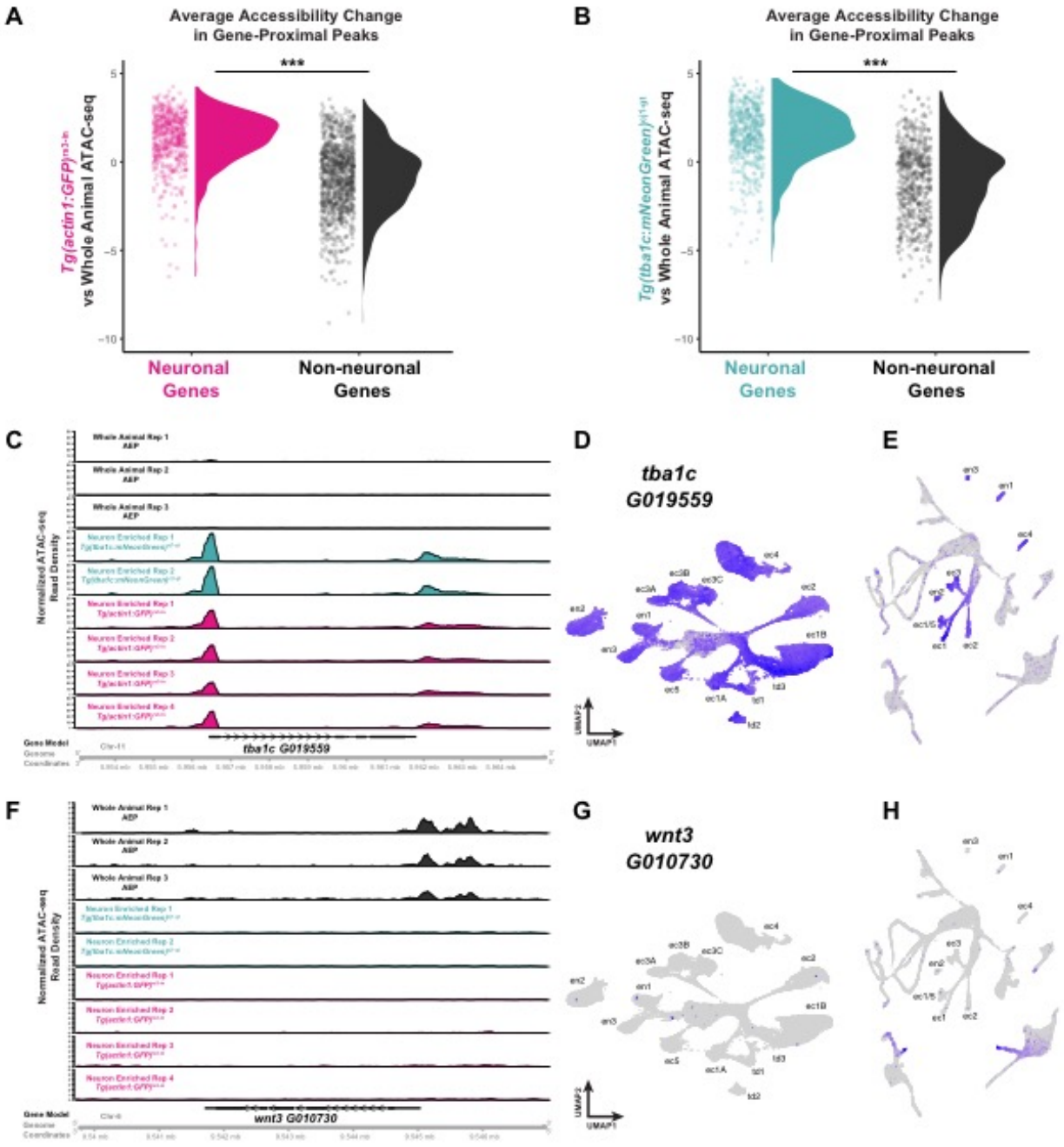


Figure 2.3. Characterization of the neuronal chromatin landscape using ATAC-seq. A total of six neuron-enriched ATAC-seq libraries were generated for this study: two from the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line and four from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ transgenic line (Keramidioti et al., 2023). Three whole animal ATAC-seq libraries (AEP1-3) were used from (Cazet et al., 2023) to test for enrichment of neuronal peaks in the neuron-enriched ATAC-seq libraries . (A-B) Average accessibility change in gene-proximal peaks in (A) *Tg(actin1:GFP)*$^{rs3\text{-}in}$ and (B) *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ as compared to whole animal AEP data sets. In both comparisons, there is significant enrichment ($p < 0.001$, t-test) of peaks within 10,000 base pairs of the transcription start site of neuronal genes as compared to non-neuronal

genes. (C-E) Example (C) ATAC-seq data tracks and scRNA-seq expression data for the *alpha tubulin* gene (G019559) in (D) the neural UMAP (this study) and (E) in the previously published whole animal data set (Cazet et al., 2023; Siebert et al., 2019). (D,E) *Alpha tubulin* has enriched expression in all neurons as compared to non-neuronal cell types. (C) Consistent with this, we observed ATAC-seq peaks enriched in the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ (teal) and the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (magenta) libraries as compared to the whole animal libraries (black). (F-H) Example (F) ATAC-seq data tracks and (G,H) scRNA-seq expression data for the *wnt3* gene (G010730). *Wnt3* expression is largely restricted to epithelial cells at the oral end (Hobmayer et al., 2000) and is absent from neuronal cell types. Consistent with this, we observed ATAC-seq peaks enriched in the whole animal libraries (black) as compared to the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ (teal) and *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (magenta) libraries.

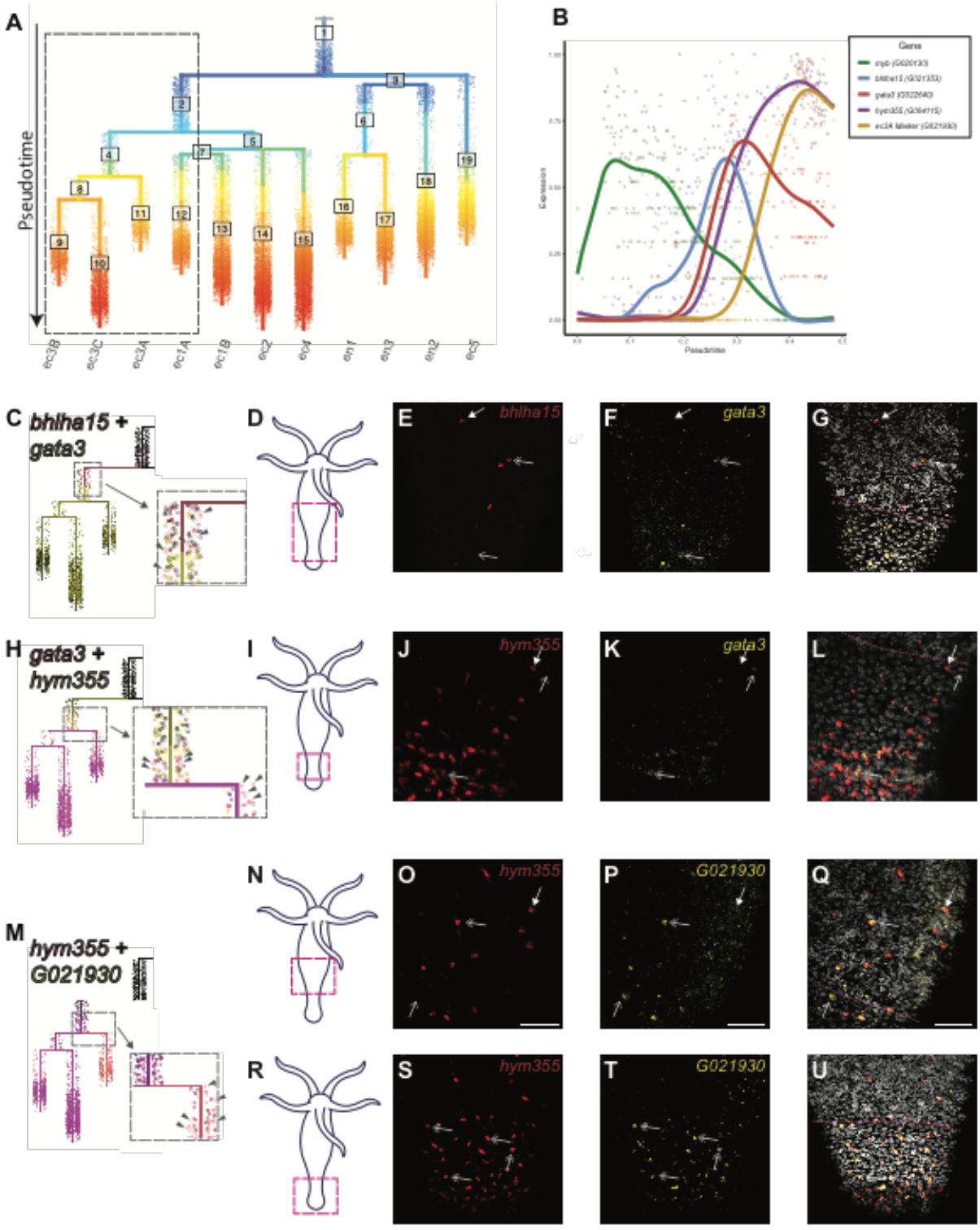


Figure 2.4. Trajectory reconstruction of *Hydra* neurogenesis indicates that ectodermal and endodermal neurons follow unique development pathways. (A) Differentiation trajectories describing the fate specification of *Hydra's* 11 neuron subtypes were reconstructed from the single cell data using URD (Farrell et al., 2018). Interstitial stem cells were selected as the "root"

or starting point of the tree and each neuron subtype was selected as a "tip", or end of the tree. Tree is colored by pseudotime (developmental time), with earlier pseudotime at the top of the tree and later pseudotime at the bottom of the tree. Boxed area represents differentiation pathway validated in B-U. (B) Spline plot representation of five genes expressed during ec3A differentiation. Pseudotime is depicted on the X axis, with earlier pseudotime on the left and later pseudotime on the right. Gene expression levels are depicted on the Y axis. Each dot on the graph represents the average gene expression of 5 cells. (C-U) Validation of predicted transition states during differentiation of the ec3 subtypes using double fluorescent RNA *in situ* hybridization (FISH). Gene expression states in FISH images are indicated by the following arrow types: cells expressing only the first gene are depicted with a closed arrow, cells expressing only the second gene are depicted with an open arrow, and cells co-expressing both genes are depicted with a double arrow. (C-G) Validation of early ec3 differentiation transition states co-expressing *bhlha15* and *gata3*. (C) *bhlha15* (magenta) and *gata3* (yellow) gene expression visualized on the URD differentiation trajectory. Cells co-expressing genes are shown in orange and are indicated on the magnified section with arrow heads. Cells not expressing either gene are shown in black. (D-G) Confocal microscopy of double FISH. (D) Area on *Hydra* body imaged. Double FISH to identify (E) *bhlha15* (red) and (F) *gata3* (yellow) expressing cells. (G) The overlay shows nuclei labeled with Hoechst (gray). (H-L) Validation of mid ec3 differentiation transition states co-expressing *hym355* and *gata3*. (H) *hym355* (magenta) and *gata3* (yellow) gene expression visualized on the URD differentiation trajectory. Cells co-expressing genes are shown in orange and are indicated on the magnified section with arrow heads. Cells not expressing either gene are shown in black. (I-L) Confocal microscopy of double FISH. (I) Area on *Hydra* body imaged. Double FISH to identify (J) *hym355* (red) and (K) *gata3* (yellow) expressing cells. (L) The overlay shows nuclei labeled with Hoechst (gray). (M-U) Validation of late ec3 differentiation transition states co-expressing *hym355* and ec3A marker *G021930*. (M) *hym355* (magenta) and ec3A marker *G021930* (yellow) gene expression visualized on the URD differentiation trajectory. Cells co-expressing genes are shown in orange and are indicated on the magnified section with arrow heads. Cells not expressing either gene are shown in black. (N-U) Confocal microscopy of double FISH. (N, R) Areas on *Hydra* body imaged (O, S) *hym355* (red) and (P, T) ec3A marker (yellow) expressing cells. (Q, U) The overlay shows nuclei labeled with Hoechst (gray). Scale bar: 50 μm. Pink dotted line in microscopy images indicates the border between the body column and peduncle as determined by nuclei morphology.

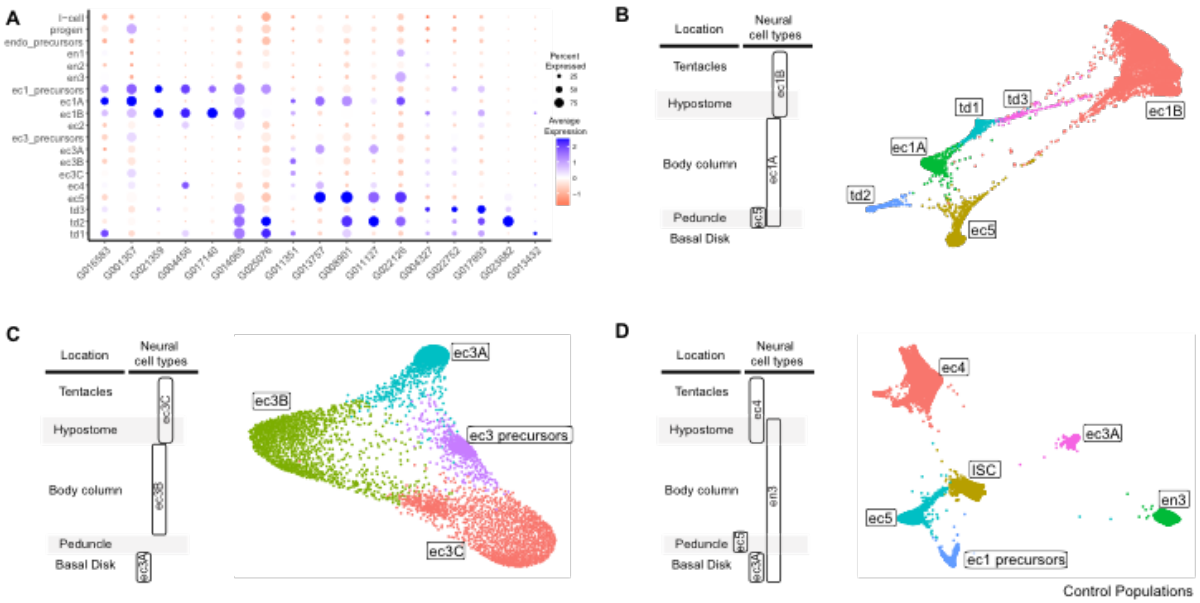


Figure 2.5. Force Directed Layout (FDL) demonstrates evidence of transdifferentiation between neuron subtypes. The neuron subtypes that could plausibly transdifferentiate were isolated and analyzed using FDL (Farrell et al., 2018; Fruchterman and Reingold, 1991). (A) Dotplot showing co-expression of markers between ec1A, ec1B, ec5, and td1-3. (B) FDL of subtypes ec1A, ec1B, ec5, td1, td2, td3 showing that ec1A (body column) connects to ec1B (hypostome/tentacles) through two intermediate transdifferentiation populations (td1 and td3). ec1A also appears to give rise to ec5 (peduncle) both directly and through an intermediate td2 population. (C) FDL of subtypes ec3A, ec3B, ec3C and ec3_precursors. ec3A (basal disk) and ec3C (hypostome/tentacles) appear to differentiate from ec3 precursors as well as transdifferentiate from ec3B (body column) neurons. (D) FDL control using populations (ec4, ec5, ec1 precursors, I-cell, ec3A, en3) with little to no predicted transcriptional similarity. Populations with more similar transcriptional profiles (I-cell, ec5, ec1 precursors) had stronger connections while populations with no similarity (ec4, ec3A, en3) had no connections.

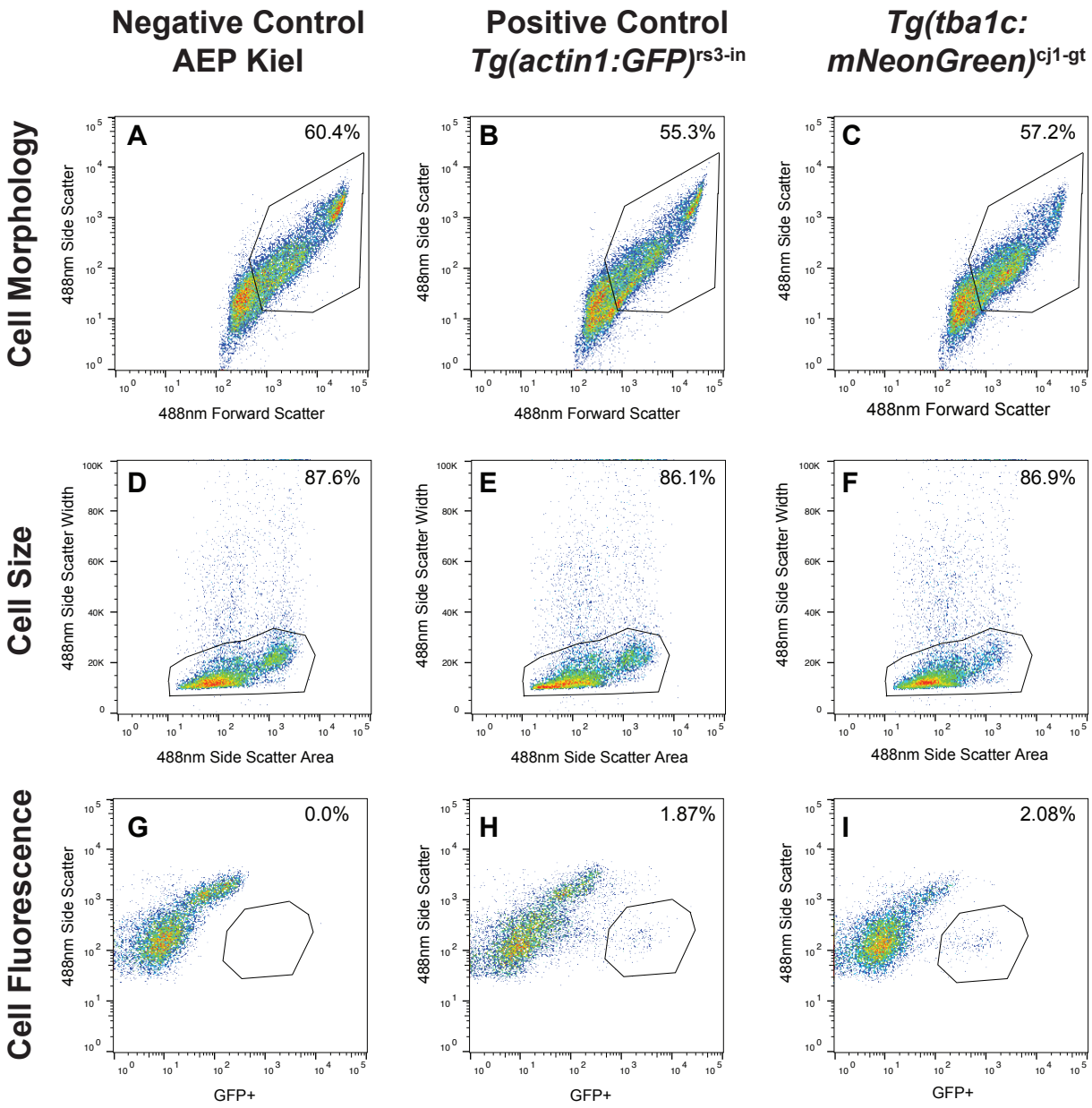


Figure 2.6. Gating for Fluorescent Activated Cell Sorting (FACS) of transgenic *Hydra* cells. Gating specifications implemented during collection of transgenic *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ *Hydra* cells used in single cell RNA sequencing and ATAC-seq. *Tg(actin1:GFP)*$^{rs3\text{-}in}$ cells were collected using gates previously applied in (Siebert et al., 2019) and were used as a positive control to guide gating for this study. Gates were set to account for cell morphology (A-C), exclusion of doublets based on cell size (D-F), and GFP expression based on cell fluorescence (G-I). Representative gates for collection of the negative FACS control AEP Kiel (A, D, G), positive FACS control *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (B, E, H) (Keramidioti et al., 2023), and *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ cells (C, F, I), with the percent of total cells within each gate shown in the top right corner of each panel.

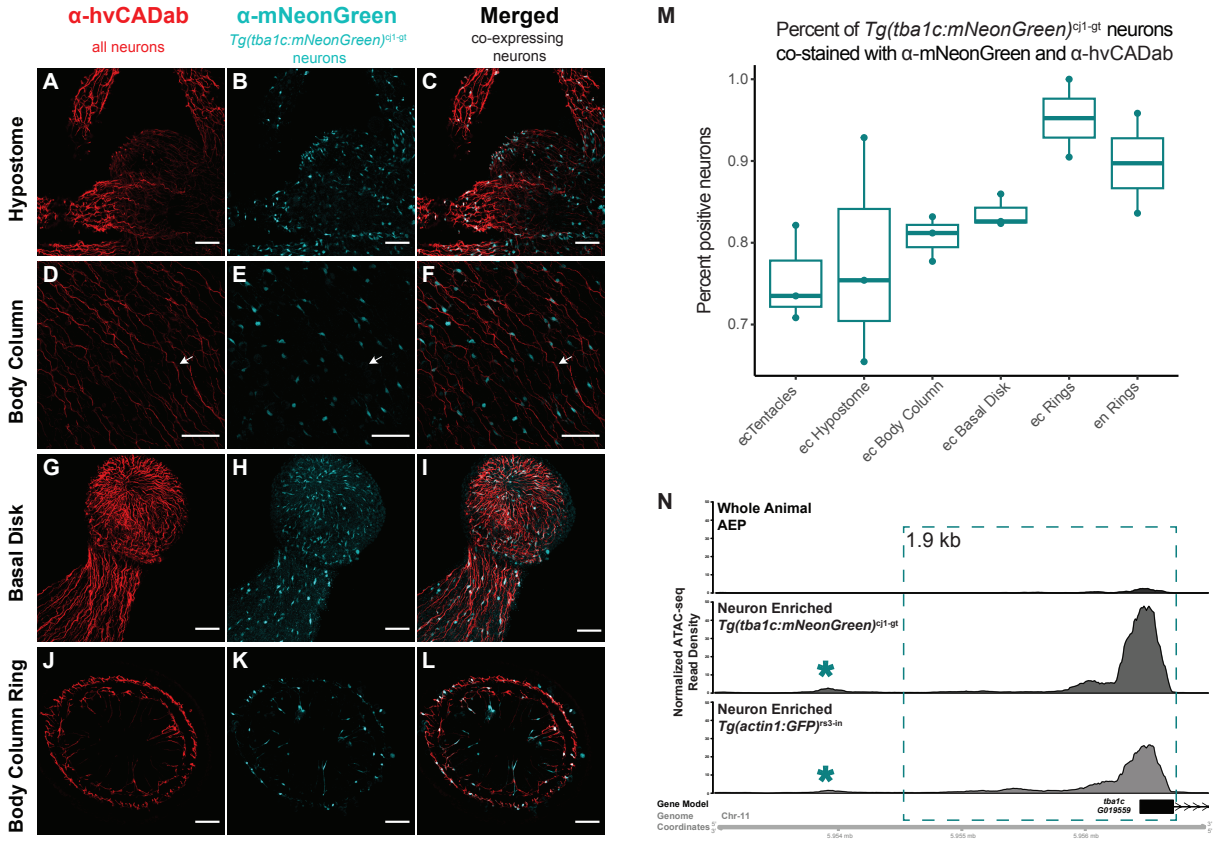


Figure 2.7. *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ line shows mosaic transgene expression despite germline transmission. (A-L) Co-localization of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ nerve cells with the hvCADab immunostained nerve net; hvCADab labels all neurons (Keramidioti et al., 2023). Confocal images show different positions along the body column of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ *Hydra*: (A-C) hypostome and tentacles, (D-F) ectoderm in the body column, (G-I) basal disk and (J-L) ring of tissue excised from the gastric region perpendicular to the oral/aboral axis. The nerve net is immunostained with hvCADab and shown in red; the fluorescent signal (mNeonGreen) of the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ nerve cells has been false-colored with cyan to increase contrast. All transgenic neurons are hvCADab positive. However, some hvCADab stained nerve cells have a very weak or no mNeonGreen fluorescence (white arrow). The nerve cell body can be clearly identified by DAPI staining (not shown). Scale bar: 50 µm. (M) Quantification of confocal images calculating the percent of *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ cells co-stained with α-mNeonGreen and the pan-neuronal α-hvCADab across different sections of the body. n ≥3 animals used for each body section with 1-4 stacks counted per section. ec, ectodermal; en, endodermal (N) Regions of open chromatin (peaks) upstream of the *tba1c* gene. When making this line, we used peaks from the whole animal 105 ATAC-seq data (Siebert et al., 2019) as a guide for identifying the regulatory region, and cloned 1901 bp upstream of the *tba1c* transcription start site (TSS), as indicated by dashed box. Our new neuron-enriched ATAC-seq data identifies an additional peak ~3,000 bp upstream of the TSS that was not captured in our cloning process, indicated by asterisks.

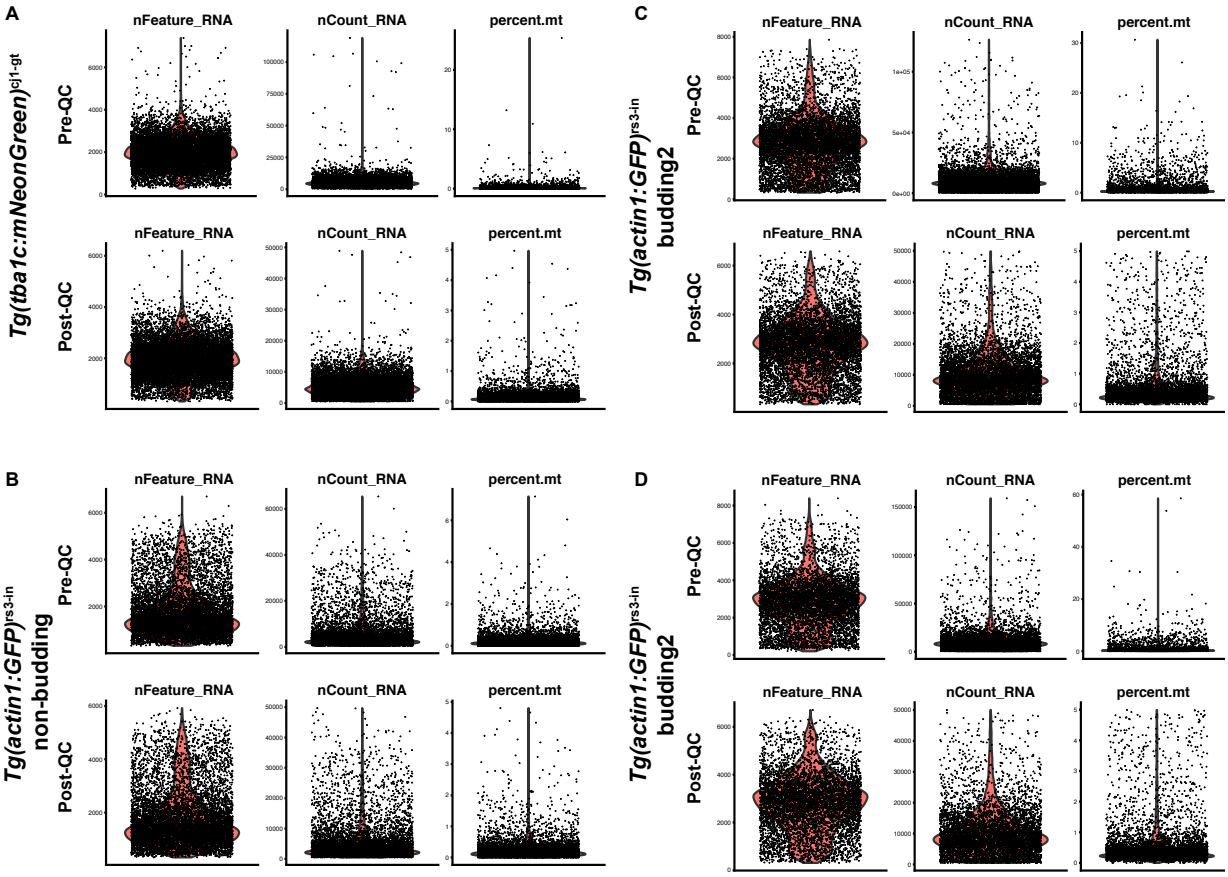


Figure 2.8. Quality control plots for single cell RNA-seq libraries. The pre- and post-quality control (QC) features are shown for each single cell RNA-seq library used in this study. Each dot is a single cell transcriptome from the respective library and the violin plots show the distribution of the data over the following parameters: nFeature_RNA refers to the number of genes detected in each single cell transcriptome, nCount_RNA refers to the number of transcripts detected in each single cell transcriptome, and percent.mt refers to the percentage of mitochondrial transcripts detected in each single cell transcriptome. The cells chosen for downstream analysis fell within the following parameters: 1) 300 - 7,000 uniquely expressed genes, 2) 500 - 50,000 transcripts, and 3) less than 5% mitochondrial reads. (A) QC plots for the neuron enriched library derived from *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic *Hydra*. (B) QC plots for the neuron enriched library derived from non-budding *Tg(actin1:GFP)*$^{rs3\text{-}in}$ *Hydra* (Keramidioti et al., 2023). (C) QC plots for the first neuron enriched library derived from budding *Tg(actin1:GFP)*$^{rs3\text{-}in}$ *Hydra*. (D) QC plots for the second neuron enriched library derived from budding *Tg(actin1:GFP)*$^{rs3\text{-}in}$ *Hydra.*

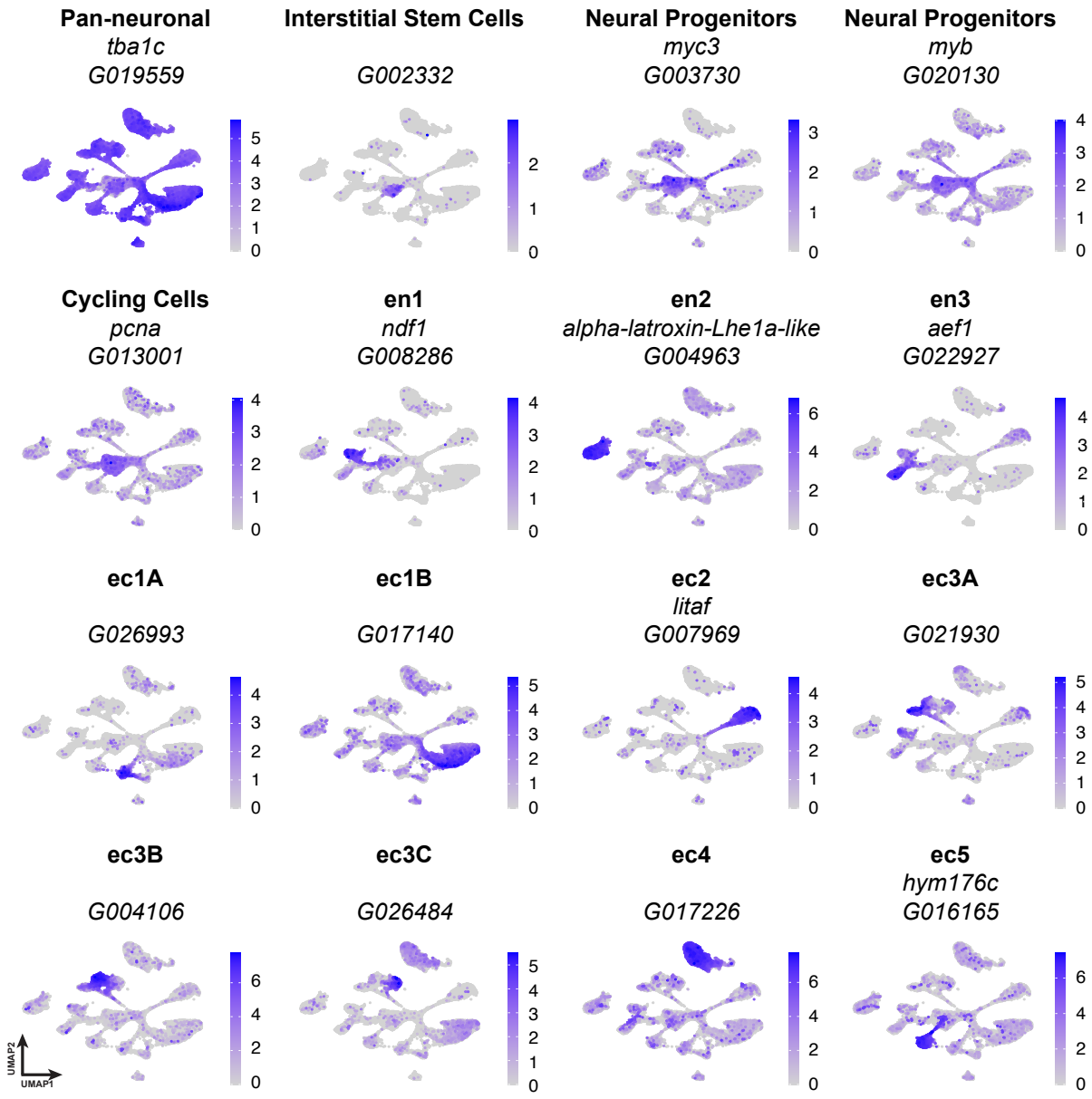


Figure 2.9. Cluster annotation using known cell-type markers. The color scale for each plot represents the log normalized, scaled UMI counts for the indicated gene. Markers used were identified in our previous study (Siebert et al., 2019). The title of each plot indicates the cell type annotated (in bold), the gene or protein name below if applicable, and the geneID from the *Hydra vulgaris* AEP genome (Cazet et al., 2023).

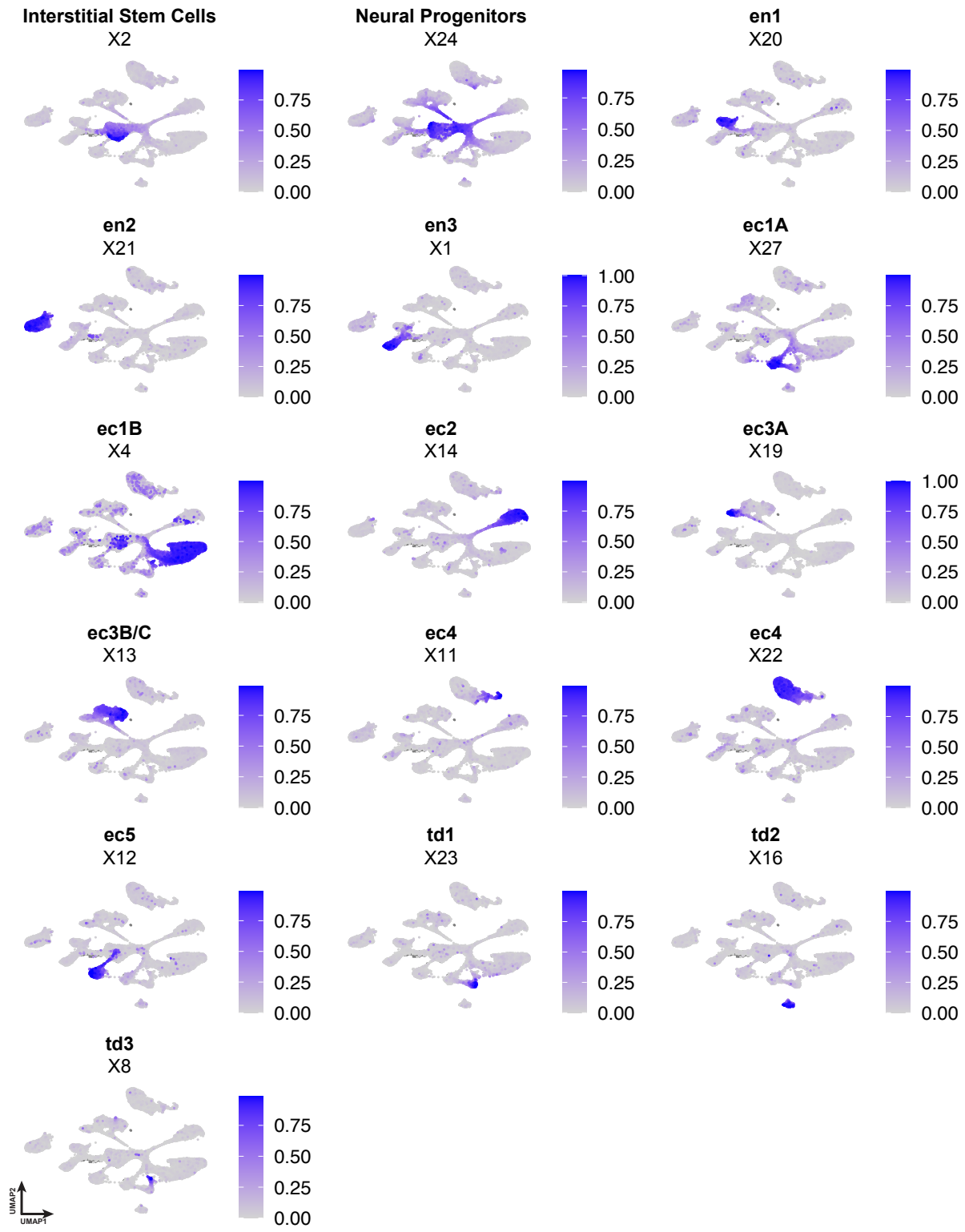


Figure 2.10. Selected metagenes identified in NMF analysis. Metagenes are groups of co-expressed genes as identified by non-negative matrix factorization (NMF) (Kotliar et al., 2019). Known gene expression patterns were used to annotate NMF gene modules. Inappropriate co-expression of metagenes was used in URD to identify cells likely to be doublets.

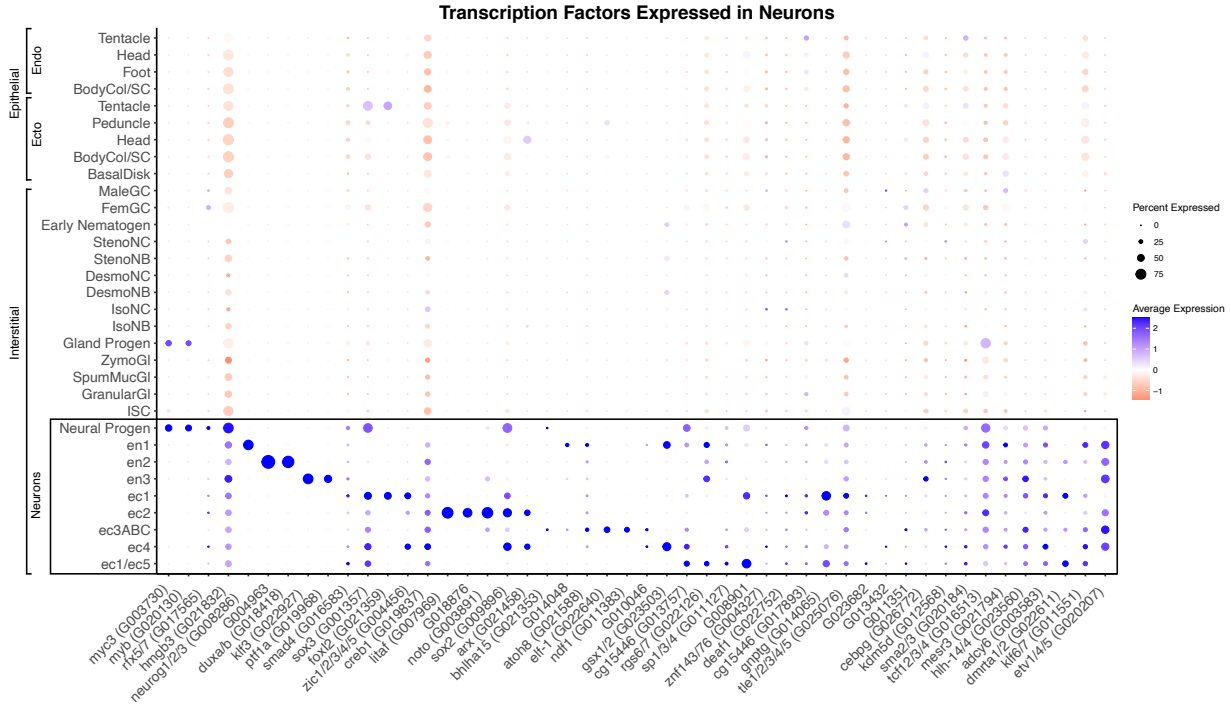


Figure 2.11. Forty-eight putative transcription factors (TFs) are uniquely expressed in *Hydra* neurons. Dotplot representation of putative TFs expressed in *Hydra* neurons shown across all *Hydra* cell types. Single cell data are from (Siebert et al., 2019) remapped to the *Hydra vulgaris* AEP genome (Cazet et al., 2023) and do not include data collected for this study. Gene expression values are shown via average expression across all cell types (dot color) and percent expression within each cluster (dot size), with the largest and darkest dots having the highest gene expression. The first three TFs are expressed in progenitors and the remaining putative TFs are expressed in one or more neuron subtypes. Neurons and neuronal progenitors are indicated by the black box. Endo/en: endodermal, ecto/ec: ectodermal, SC: stem cell, GC: germ cell, NC: nematocyst, NB: nematoblast, Gl: gland cell, ISC: interstitial stem cell.

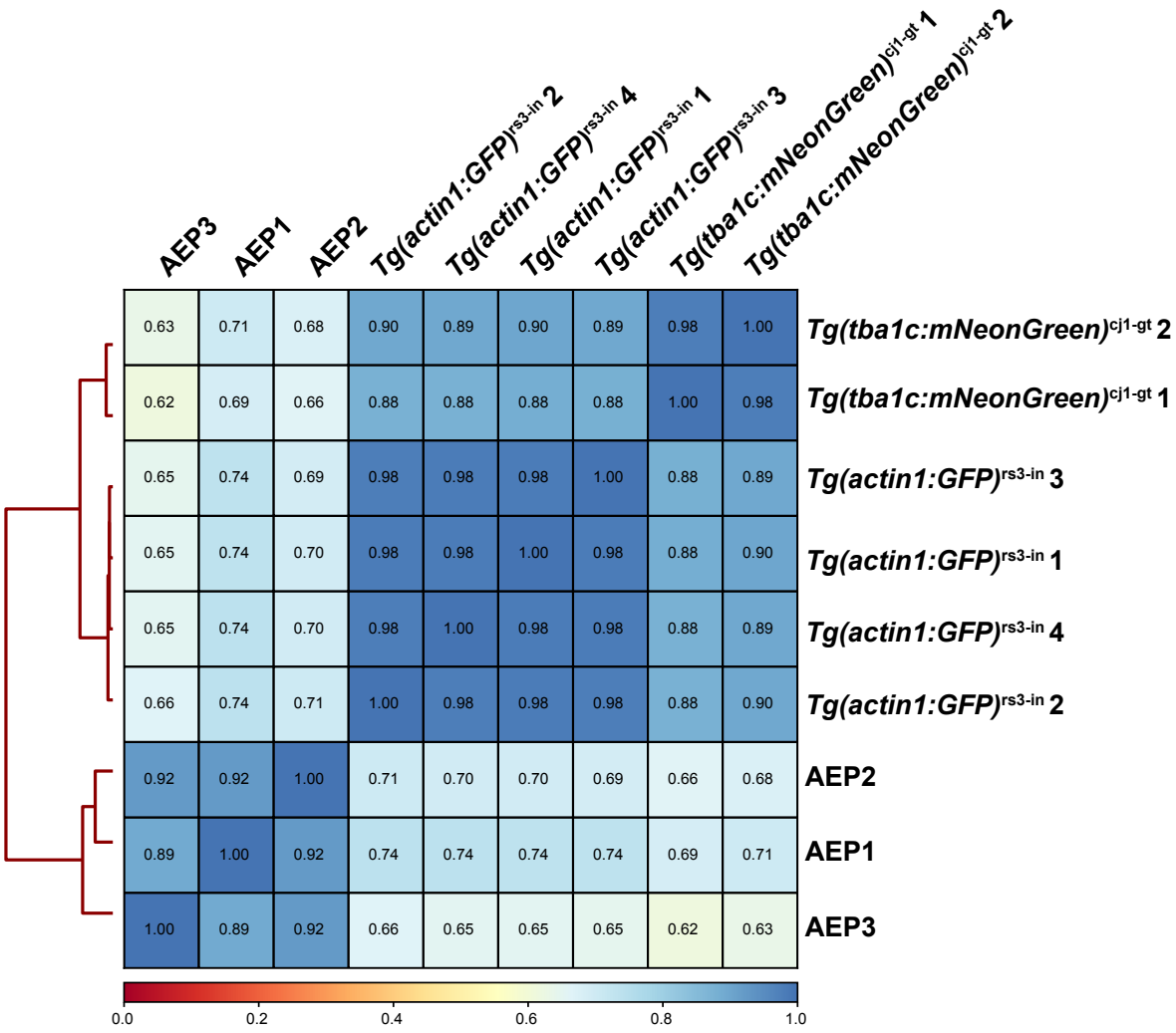


Figure 2.12. Neuron-enriched ATAC-seq libraries show high reproducibility. A correlation plot for all ATAC-seq samples. Calculations using a Spearman correlation show that all biological replicates are nearly identical. A total of 6 neuron-enriched ATAC-seq libraries were generated for this study: two from the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line and four from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ transgenic line (Keramidioti et al., 2023). Three whole animal ATAC-seq libraries (AEP1-3) were used from (Cazet et al., 2023) as controls and as benchmarks for data standards.

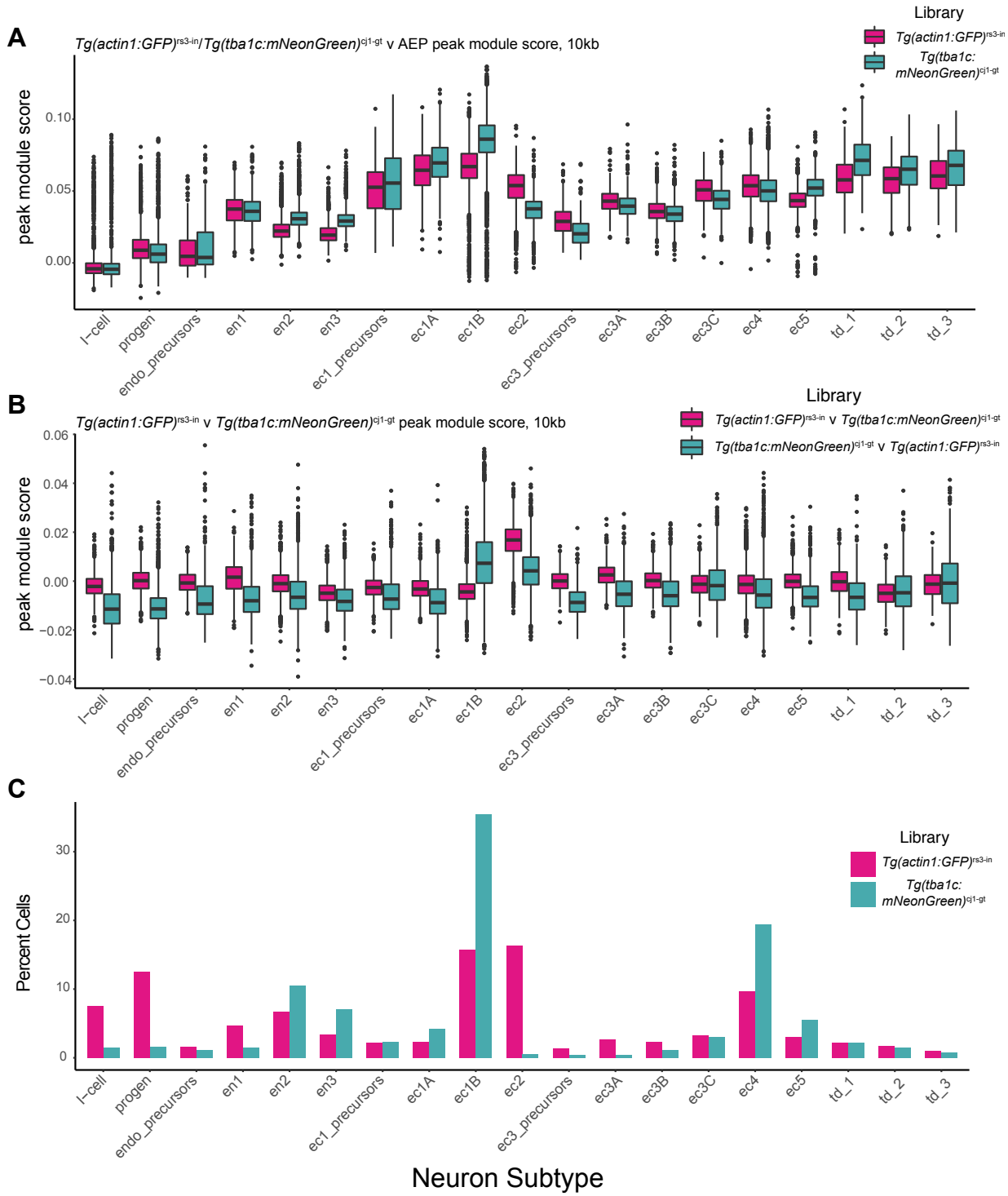


Figure 2.13. Approximation of neural subtypes sequenced in neuron-enriched, bulk ATAC-seq libraries. To identify the subtypes we captured and sequenced from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ (Keramidioti et al., 2023) and *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic lines, we identified genes associated with the top differentially expressed peaks (over 3 logFC) as compared to the non-neuron enriched *Hydra vulgaris* strain AEP libraries. We used the differentially expressed peaks found within 10kb of the nearest gene to assign module scores to

the scRNA-seq data as an approximation of cell type enrichment. These are compared to the scRNA-seq data. (A) Overall enrichment of all neuron subtypes in the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ and *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ ATAC-seq libraries as compared to the AEP whole animal ATAC-seq libraries (nGreen vs AEP, *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ vs AEP). Neuron subtypes are on the x-axis and the peak module score is on the y-axis. Scores over 0 are considered enriched. (B) Overall enrichment or depletion of neuron subtypes in the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ and *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ ATAC-seq libraries as compared to each other (nGreen vs *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$, *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ vs *Tg(actin1:GFP)*$^{rs3\text{-}in}$). Neuron subtypes are on the x-axis and the peak module score is on the y-axis. Scores over 0 are considered enriched. (C) The percent of cells of each subtype as compared to all cells captured in each *Tg(actin1:GFP)*$^{rs3\text{-}in}$ and *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ scRNA seq library. Neuron subtype is on the x-axis and the percent of cells out of the total cells sequenced is on the y-axis.

Library	Animals Sorted	Days of Starvation	FACS Collection Time (min)	# GFP+ Cells Collected	# Cells Sequenced
Tg(tba1c: mNeonGreen)$^{cj1\text{-}gt}$	90	3	42	78,122	11,138
Tg(actin1:GFP)$^{rs3\text{-}in}$ non-budding	80	5	23	138,586	7,988
Tg(actin1:GFP)$^{rs3\text{-}in}$ budding1	70	3	23	100,429	7,374
Tg(actin1:GFP)$^{rs3\text{-}in}$ budding2	70	3	23	100,429	6,124

Table 2.1. Information on scRNA-seq library preparation. A total of 4 neuron-enriched scRNA-seq libraries were generated for this study: one from the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line and three from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ transgenic line (Keramidioti et al., 2023). A description of column labels are as follows: "Library" refers to the sample type and biological replicate. "Animals Sorted" refers to the number of animals dissociated and sorted using FACS. "Days of Starvation" refers to the days animals were starved prior to dissociation and sorting. "FACS Collection Time (min)" refers to the number of minutes FACS sorting took to collect the number of cells reported in "GFP+ Cells Collected". "# Cells Sequenced" refers to the total number of cells captured and sequenced per library using the Chromium Next GEM Single Cell 3' kit v3.1 (10x Genomics).

Library	Data From	Sample Type	Sequencing Method	# Cells Sequenced	# Cells Used in Study	Median UMI	Median Gene
Tg(tba1c: mNeonGreen)$^{cj1\text{-}gt}$	This study	Neuron Enriched	10x Genomics	11138	10971	4941	1941
Tg(actin1:GFP)$^{rs3\text{-}in}$ nonbudding	This study	Neuron Enriched	10x Genomics	7988	7204	2928	1488
Tg(actin1:GFP)$^{rs3\text{-}in}$ budding1	This study	Neuron Enriched	10x Genomics	7374	6313	8820	2834
Tg(actin1:GFP)$^{rs3\text{-}in}$ budding2	This study	Neuron Enriched	10x Genomics	6124	5183	8982	2904
D12_N2	Siebert *et al.* 2019	Neuron Enriched	Drop-seq	1953	1575	917	513
D12_N2	Siebert *et al.* 2019	Neuron Enriched	Drop-seq	1264	1007	912	531
D01_P2	Siebert *et al.* 2019	Whole Animal	Drop-seq	1256	371	2640	1255
D02_P1	Siebert *et al.* 2019	Whole Animal	Drop-seq	3343	320	4172	1724
D02_CO	Siebert *et al.* 2019	Whole Animal	Drop-seq	2329	313	2249	1101
D03_K1	Siebert *et al.* 2019	Whole Animal	Drop-seq	1958	287	4280	1503
D03_FM	Siebert *et al.* 2019	Whole Animal	Drop-seq	886	276	2847	1154
D06_KI	Siebert *et al.* 2019	Whole Animal	Drop-seq	1937	267	2173	1016
D01_D1	Siebert *et al.* 2019	Whole Animal	Drop-seq	1023	216	4701	1818
D02_PB	Siebert *et al.* 2019	Whole Animal	Drop-seq	1562	186	3288	1516
D06_MA	Siebert *et al.* 2019	Whole Animal	Drop-seq	384	172	2572	1194
D11_BU	Siebert *et al.* 2019	Whole Animal	Drop-seq	3207	169	1380	807
D03_MA	Siebert *et al.* 2019	Whole Animal	Drop-seq	702	136	4933	1766
D11_PO	Siebert *et al.* 2019	Whole Animal	Drop-seq	2058	105	1565	781

Table 2.2. Metrics for single cell libraries used in this study. 18 single cell RNA-seq libraries were used in this study: 4 neuron-enriched libraries were generated for this study and 14 libraries from (Siebert et al., 2019) were used in addition. A description of the column labels are as follows: "Library" refers to the sample type and biological replicate. "Data From" indicates whether the library was generated for this study or for Siebert *et al*. (2019). "Sequencing Method" refers to the scRNA-seq collection method. "# Cells Sequenced" indicates the total number of cells in each library prior to filtering and/or subsetting. "# Cells Used in Study" refers to the number of cells in each library after filtering and/or subsetting. "Median UMI" refers to the median number of unique molecular identifiers (i.e. transcripts) in each library, and "Median Gene" refers to the median number of unique genes in each library.

Cell Type	# Cells	Median UMI	Median Gene
ISC	1575	12946	3182
progenitors	3221	9101	3058
endo precursors	456	8741	2641
en1	1463	2648	1391
en2	2621	3702	1654
en3	1545	5547	2117
ec1 precursors	787	6437	2484
ec1A	1278	3646	1666
ec1B	7134	4938	1993
ec2	3347	3984	1789
ec3 precursors	461	10221	2981
ec3A	730	5434	2104
ec3B	1280	7548	2480
ec3C	1371	6319	2242
ec4	4632	3428	1466
ec5	1505	8332	2372
td1	757	4453	1852
td2	570	3792	1782
td3	338	4492	1944

Table 2.3. Median genes and UMIs per cell per state. Nineteen transcriptionally distinct clusters were annotated. A description of the column labels are as follows: “Cell Type” indicates the cell type. “# Cells” indicates how many cells were sequenced from each subtype. “Median UMI” refers to the median number of unique molecular identifiers (i.e. transcripts) from cells in each cell type, and “Median Gene” refers to the median number of unique genes in each cell type.

Library	Animals Sorted	Days of Starvation	FACS Collection Time (min)	GFP+ Cells Collected	Tagmentation Protocol Time (min)	Total PCR Cycles for Amplification
Tg(tba1c: mNeonGreen)$^{cj1\text{-}gt}$ 1	95	3	19	56,250	70	11
Tg(tba1c: mNeonGreen)$^{cj1\text{-}gt}$ 2	95	5	22	60,149	59	12
Tg(actin1:GFP)$^{rs3\text{-}in}$ 1	70	2	29	60,065	58	11
Tg(actin1:GFP)$^{rs3\text{-}in}$ 2	70	3	31	57,433	53	11
Tg(actin1:GFP)$^{rs3\text{-}in}$ 3	70	3	40	60,188	54	10
Tg(actin1:GFP)$^{rs3\text{-}in}$ 4	70	2	24	62,495	56	10

Table 2.4. Information on ATAC-seq library preparation. A total of 6 neuron-enriched ATAC-seq libraries were generated for this study: two from the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line and four from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ transgenic line (Keramidioti et al., 2023). A description of column labels are as follows: "Library" refers to the sample type and biological replicate. "Animals Sorted" refers to the number of animals dissociated and sorted using FACS. "Days of Starvation" refers to the days animals were starved prior to dissociation and sorting. "FACS Collection Time (min)" refers to the number of minutes FACS sorting took to collect the number of cells reported in "GFP+ Cells Collected". "Tagmentation Protocol Time (min)" refers to the total length of the ATAC-seq tagmentation protocol (Buenrostro et al., 2013; Corces et al., 2017), and "Total PCR Cycles for Amplification" refers to how many PCR cycles each library underwent during initial amplification.

Library	Sample Type	Library Origin	Total Reads	Final Mapped Read Pairs	Alignment Rate	TSS Enrichment	Reproducible Peaks	Self-consistency Ratio	Rescue Ratio
AEP1	whole animal	Cazet *et al.* 2023	44373180	18530668	85.83%	5.41	50151	1.009	1.261
AEP2	whole animal	Cazet *et al.* 2023	38095986	16220852	79.09%	4.09			
AEP3	whole animal	Cazet *et al.* 2023	30846609	16788076	85.05%	3.74			
Tg(tba1c:mNeonGreen)$^{cj1\text{-}gt}$ 1	Neuron Enriched	This study	97611928	27969643	93.69%	4.40	46528	1.092	1.086
Tg(tba1c:mNeonGreen)$^{cj1\text{-}gt}$ 2	Neuron Enriched	This study	97349346	27327438	89.52%	4.64			
Tg(actin1:GFP)$^{rs3\text{-}in}$ 1	Neuron Enriched	This study	108115128	36065847	90.25%	4.30	50389	1.087	1.784
Tg(actin1:GFP)$^{rs3\text{-}in}$ 2	Neuron Enriched	This study	113433254	32885448	89.10%	4.22			
Tg(actin1:GFP)$^{rs3\text{-}in}$ 3	Neuron Enriched	This study	107557708	37807471	94.08%	4.02			
Tg(actin1:GFP)$^{rs3\text{-}in}$ 4	Neuron Enriched	This study	115038941	33599627	91.90%	4.51			

Table 2.5. Determining quality of neuron-enriched ATAC-seq libraries. A total of 6 neuron-enriched ATAC-seq libraries were generated for this study: two from the *Tg(tba1c:mNeonGreen)*$^{cj1\text{-}gt}$ transgenic line and four from the *Tg(actin1:GFP)*$^{rs3\text{-}in}$ transgenic line (Keramidioti et al., 2023). Three whole animal ATAC-seq libraries (AEP1-3) were used from (Cazet et al., 2023) as controls and as benchmarks for data standards. A description of the column labels are as follows: "Library" refers to the sample type and biological replicate. "Data From" indicates whether the library was generated for this study or for Cazet *et al*., 2023 "Total Read Pairs" refers to the number of raw reads generated for each library. "Final Mapped Read Pairs" refers to the number of read pairs remaining after duplicated, unmapped, and ambiguously mapped reads were removed. "Alignment Rate" refers to the percentage of final mapped read pairs aligned to the reference genome out of the total number of read pairs. ENCODE considers alignment rates over 80% to be acceptable (encodeproject.org/atac-seq) (Landt et al., 2012). "Transcription Start Site (TSS) Enrichment" refers to the fold enrichment in ATAC-seq signal at the TSS relative to regions flanking the TSS by +/- 1 kb. ENCODE considers TSS scores >5 to be acceptable, but these scores are highly dependent on the model system used and may be reflective of the native biology of the system. "Reproducible Peaks" refers to the number of peaks within each library that were biologically reproducible in at least 2 pairwise comparisons using an irreproducible discovery rate cutoff of 0.1. ENCODE considers the number of reproducible peaks >50,000 to be acceptable. The self-Consistency Ratio" measures reproducible peak consistency within a single dataset and the "Rescue Ratio" measures consistency between datasets. ENCODE considers self-consistency ratios and rescue ratios <2 to be ideal.

Supplemental tables and data can be found at DOI: 10.1101/2023.03.15.531610

Appendix A: Relationship between neural activity and neuronal cell fate in regenerating *Hydra* revealed by cell-type specific imaging

This chapter was originally published as a pre-print in *BioRxiv*:

Escobar A, K Soonyoung, AS Primack, G Duret, CE Juliano, JT Robinson. 2023. Relationship between neural activity and neuronal cell fate in regenerating Hydra revealed by cell-type specific imaging. bioRxiv *2023.03.19.533365*

I made the following contributions to the work presented in appendix A: I identified the pan-neuronal promoter used for the transgenic line *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$ created for this study, generated the transgenic animals, wrote methods, and provided feedback on the manuscript.

ABSTRACT

Understanding how neural circuits are regenerated following injury is a fundamental question in neuroscience. *Hydra* is a powerful model for studying this process because it has significant and reproducible regenerative abilities, a simple and transparent body that allows for whole nervous system imaging, and established methods for creating transgenics with cell-type-specific expression. In addition, cnidarians such as *Hydra* split from bilaterians (the group that encompasses most model organisms used in neuroscience) over 500 million years ago, so similarities with other models likely indicates deeply conserved biological processes. *Hydra* is a long-standing regeneration model and is an emerging model for neuroscience; however, relatively little is known regarding the restoration of neural activity and behavior following significant injury. In this study, we ask if regenerating neurons reach a terminal cell fate and then reform functional neural circuits, or if neural circuits regenerate first and then guide the constituent cells toward their terminal fate. To address this question, we developed a dual-expression transgenic *Hydra* line that expresses a cell-type-specific red fluorescent protein

(tdTomato) in ec5 peduncle neurons, and a calcium indicator (GCaMP7s) in all neurons. This transgenic line allowed us to monitor neural activity while we simultaneously track the reappearance of terminally differentiated ec5 neurons as determined by the expression of tdTomato. Using SCAPE (Swept Confocally Aligned Planar Excitation) microscopy, we tracked both calcium activity and expression of tdTomato-positive neurons in 3D with single-cell resolution during regeneration of *Hydra's* aboral end. We observed tdTomato expression in ec5 neurons approximately four hours before the neural activity begins to display synchronized patterns associated with a regenerated neural circuit. These data suggest that regenerating neurons undergo terminal differentiation prior to re-establishing their functional role in the nervous system. The combination of dynamic imaging of neural activity and gene expression during regeneration make *Hydra* a powerful model system for understanding the key molecular and functional processes involved in neuro-regeneration following injury.

INTRODUCTION

Neural regeneration capacity is widely exemplified across animals. The extent of these regenerative capacities ranges from cellular regeneration to whole body reformation, with some of the most extreme examples of nervous system regeneration found in flatworms (e.g., *Schmidtea mediterranea)* (Lobo et al., 2012; Reddien, 2018), cnidarians (e.g., *Hydra vulgaris*) (Reddy et al., 2019; Sarras Jr, 2019; Vogg et al., 2019b), the replacement of complete innervated limbs in salamanders (e.g., *Ambystoma mexicanum*) (Vieira et al., 2019; Wells et al., 2021) or spinal cord injury recovery in zebrafish (e.g., *Danio rerio*) (Beffagna, 2019; Vandestadt et al., 2021). By contrast, mammals exhibit limited regenerative abilities, along with a complex immune response that slows neural regrowth (Julier et al., 2017). Understanding the molecular

and cellular mechanisms that drive nervous system regeneration in highly regenerative animals will likely inform the development of neural repair therapies for humans (Lu et al., 2009). In particular, we need to understand how newly regenerated neurons rebuild functional neural circuits. The cnidarian polyp *Hydra* has a simple neural structure, extensive neuron regenerative capabilities (Vogg et al., 2019b; Wilson-Sanders, 2011), established genetic tools (Chapman et al., 2010; Wittlieb et al., 2006), and is an emerging neuroscience model (Bosch et al., 2017). These features make *Hydr*a an excellent model for interrogating the cellular dynamics of neural circuit functional regeneration.

Hydra has a simple radial body plan organized around a single oral-aboral axis of symmetry, with the hypostome and tentacles at the oral end (i.e., the head) and the peduncle and basal disk at the aboral end. The *Hydra* body is formed by two epithelial monolayers, the inner endoderm and outer ectoderm, separated by an extracellular matrix (Szymanski and Yuste, 2019). *Hydra* has two separate nerve nets, one embedded in each of the epithelial layers. Neurons run along the entire length of the body, with a higher neuron density in the hypostome (oral end) and the peduncle and basal disc (aboral end) (Dunne et al., 1985; Keramidioti et al., 2023). Important groundwork has been done to identify neural circuits associated with specific behaviors in *Hydra* (Dupre and Yuste, 2017). This aids our investigation of neural circuit regeneration because the resumption of normal behavior indicates when a neural circuit has functionally regenerated. Specifically, *Hydra* has a defined repertoire of behaviors associated with four major non overlapping neural circuits (Brette, 2012). In this study we focus on longitudinal contractions because they are the easiest movements to track and are known to correlate with the activity of the Contraction Burst (CB) circuit. The CB circuit involves neurons

that run the length of the ectodermal epithelium, including a particularly prominent group of neurons located in the peduncle at the aboral end (Dupre and Yuste, 2017).

The transcriptional state of all eleven *Hydra* neuron subtypes has been profiled using single cell RNA-sequencing (Primack et al., 2023; Reddien, 2018). Similar to findings in *C. elegans* (De Fruyt et al., 2020), the *Hydra* neurons are best defined by unique combinatorial expression of specific genes, including transcription factors and neuropeptides (Noro et al., 2019; Taylor et al., 2021). Previous work suggests that the neuron subtypes that participate in the CB circuit are defined by combinatorial expression of various paralogs of the *hym176* neuropeptide gene (*hym176A-E*) (Noro et al., 2021; Wittlieb et al., 2006). One of these neuron subtypes is the "ec5" population, which is located in the peduncle and selectively expresses the neuropeptide Hym176C (Hansen et al., 2000; Takahashi, 2021).

In addition, cnidarians such as *Hydra* split from bilaterians (the group that encompasses most model organisms used in neuroscience) over 500 million years ago, so similarities with other models likely indicates deeply conserved biological processes. *Hydra* is a long-standing regeneration model and is an emerging model for neuroscience thanks to recent studies that elucidate stem cell differentiation trajectories in *Hydra* cells, including all neuronal subtypes (Siebert et al., 2019). However, relatively little is known about the restoration of neural activity (Brette, 2012; Knoblich et al., 2019; Uhlhaas et al., 2009) and behavior following a significant injury. As part of the regeneration process of a neural circuit, two key events occur: terminal cell differentiation and the synchronization of cell activity which elucidates recovery behavior. In *Hydra*, the order of these key events during its neural regeneration process remained

unknown. In this study, we ask if regenerating neurons reach a terminal cell fate and then reform functional neural circuits, or if neural circuits regenerate first and then guide the constituent cells toward their terminal fate.

In this study, we use the regulatory region of hym176c to drive ec5-specific nuclear expression of tdTomato along with GCaMP7s (Dana et al., 2019) in all neurons. This allowed us to perform cell-type specific 3D imaging of neural activity in both uninjured and regenerating *Hydra*. We first confirmed that ec5 neurons act in the CB circuit, and then observed the regeneration of the CB circuit by tracking the neural activity and reappearance of ec5 neurons after foot amputation. We found that ec5 neurons terminally differentiate before they synchronize with neighboring neurons. This suggests that positional cues, not neural activity cues play the dominant role in guiding neuronal cell fate in this circuit. This study provides the foundational tools and conceptual framework to better understand the molecular mechanisms that underlie the regeneration of functional neural circuits.

RESULTS

Ec5 neurons are a part of the Contraction Burst (CB) Circuit

To monitor the activity of the ec5 neurons both in uninjured *Hydra* and during regeneration, we used transgenic line *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$, which was created for this study, and in which tdTomato is specifically expressed in ec5 neurons (Figure A.1). To create this line we extracted the regulatory region of the *hym176c* gene, which is expressed specifically in ec5 neurons (Figure A.1 H) and used this to drive the expression of nuclear tdTomato. To be able to track ec5 expression in the context of the activity of the entire nervous system, we

included a second expression cassette in the construct using GCaMP7s under the control of pan-neuronal promoter *tba1c (*identified in (Primack et al., 2023)). After obtaining a founding polyp with integration of the transgene in the interstitial lineage, which includes all neurons, we propagated the line through asexual reproduction, which gave us a continuous supply of this transgenic line for experimentation. As expected, in our transgenic line only neurons in the peduncle express tdTomato, indicating that we had successfully marked ec5 neurons (Figure A.1 A).

We confirmed that GCaMP7s successfully reported calcium activity by measuring fluorescence levels during contractions. As expected, based on prior calcium imaging (Badhiwala et al., 2021; Chapman et al., 2010; Vogg et al., 2019b), the neuronal activity in the peduncle showed increased calcium activity during contractions (Figure A.1 C,F, Supplemental Video S1). We also found that the tdTomato-positive neurons were a subpopulation of the neurons that were coactive during contractions, providing further evidence that ec5 neurons are part of the CB circuit (Figure A.1 B-G). These ec5 neurons composed on average 88% of the neurons in the peduncle that showed increased calcium activity during contractions (n *Hydra*=5, Figure A.1 I). The remaining tdTomato-negative peduncle neurons (Figure A.1 I,J) in the CB circuit are likely the ec1A neurons that extend from the body column into the peduncle (Wittlieb et al., 2006).

Volumetric fluorescent imaging establishes basal levels of synchronization among neurons in the CB circuit

We used the newly created transgenic line *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$ to track the activity of ec5 neurons during regeneration to determine when normal circuit activity

resumes. However, we first needed to establish the degree of synchrony displayed by a fully functional CB circuit in an uninjured animal. This would allow us to determine if the CB circuit has fully regenerated. To quantify the level of synchrony, we imaged the spontaneous calcium activity of tdTomato-positive ec5 neurons using SCAPE (Swept Confocally Aligned Planar Excitation) 2.0 microscopy (Voleti et al., 2019). This dual-color fast light sheet imaging technique achieves a volumetric frame rate of 1.3, which compared to the movement of *Hydra*, is fast enough to accurately track the calcium dynamics of ec5 neurons during contraction. Figure A.2 A-B shows multi view Maximum Intensity Projections (MIP) of the peduncle from a contracting *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$*Hydra*at different timepoints. The calcium activity of individual ec5 neurons was tracked (colored dots, Figure A.2 C) and the corresponding traces are shown in Figure A.2 D (Supplemental Video S1-3). We calculated the cross-correlation coefficient (CC) to measure synchrony in neural activity. Considering that the CC between the ec5 neurons (n neurons=29, n *Hydra* = 1) in an uninjured *Hydra* is 0.84 +/- 0.07 (mean +/- SEM), the CC values significantly below 0.84 in regenerating *Hydra* would indicate that the CB circuit has not yet fully recovered its function.

ec5 neuron differentiation precedes neural synchronization during regeneration

Having established a quantitative measure of synchrony in an uninjured CB circuit, we next used our new transgenic line to track the reappearance and activity of ec5 neurons during regeneration. The goal of these experiments was to determine when ec5 neurons differentiated relative to when the CB circuit resumed synchronized activity, which is indicative of a regenerated and synchronous neural circuit. We hypothesized two possible scenarios: (1) newly regenerating ec5 neurons would be functionally integrated into circuits and show a high degree

of synchronization before completing terminal differentiation, or (2) ec5 neurons would express differentiation markers prior to functional integration into the CB circuit. Importantly, *hym176c*is a marker of differentiated ec5 neurons, thus the appearance of tdTomato fluorescence is a proxy for the completion of differentiation.

Since ec5 neurons are located in the peduncle (Siebert et al., 2019; Zambusi and Ninkovic, 2020), we conducted foot regeneration experiments to track their reappearance. By bisecting the animal at the midpoint between the head and foot, we completely removed the ec5 neurons from the top half of the animal. In approximately 48 hours after this injury, the foot fully regenerates from the top half (Figure A.3 A). During this time, new neurons are produced from the interstitial stem cells that reside among the ectodermal epithelial cells in the body column. Over the course of regeneration we tracked the reappearance of ec5 neurons (using tdTomato expression) and the activity of the CB circuit in the peduncle (using GCaMP7s fluorescence) by taking 20 minute recordings every four hours post amputation (hpa) (n = 5 animals) to monitor circuit reformation.

In a representative *Hydra* shown in Figure A.3, we randomly selected 5 neurons from *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$ *Hydra*to evaluate synchrony. We observed unsynchronized neural activity with no tdTomato-positive neurons (Figure A.3 B, Supplemental Video S4) from 0-28 hpa. At 32 hpa, the tdTomato-positive neurons began to appear, but these neurons exhibited a low level of synchrony (Figure A.3 C, Supplemental Video S5). At the 36 hpa mark, we observed an increase in the number of tdTomato-positive neurons along with a large increase in synchronization (Figure A.3 D, Supplemental Video S6). Due to the varying

timing of the appearance of tdTomato-positive neurons across multiple animals, we defined the time point when the tdTomato-positive (ec5) neurons first reappeared as t = 0hr (n *Hydra* = 5). With this alignment, we found a critical window of four hours that separates the first detection of tdTomato-positive cells and the synchronization of the CB circuit, named the "critical time period" (Figure A.3 A). At t = -4hr, the activity of the neurons in the regenerating foot were not synchronized (Figure A.3 E, correlation coefficient = 0.172+/-0.07 (mean +/- SEM)). At t = 0hr, the synchrony of the neurons was low (correlation coefficient = 0.239+/-0.07) even though tdTomato-positive neurons appeared. The level of synchrony increased (correlation coefficient = 0.609+/-0.03) at t = +4hr where we also observed an increase in the number of tdTomato-positive neurons. However, they were less synchronized compared to the neurons in uninjured animals (0.84 +/- 0.07 (mean +/- SEM)). Together, these data support the scenario in which ec5 neurons fully differentiate before functional integration into the CB circuit.

DISCUSSION

Although specification of neurons and assembly of neural circuits during development is relatively well studied, these processes are not as well understood during regeneration. Several research organisms, including zebrafish (*Danio rerio)*, axolotl (*Ambystoma mexicanum*), xenopus tadpoles, and planarians (*Schmidtea mediterranea)*, can regenerate large portions of their bodies including innervation and the restoration of behavior (Lee-Liu et al., 2017; Lust et al., 2022; Ross et al., 2017; Wagner et al., 2018; Zambusi and Ninkovic, 2020). These animals have provided several interesting insights, including the ability of zebrafish (Vandestadt et al., 2021) and xenopus tadpoles (Kakebeen et al., 2020) to restore neural circuit activity and behavior before regeneration is complete. *Hydra* has a unique combination of advantages as

compared to existing neuronal regeneration models which allows us to image the activity of the entire nervous system at single cell resolution. In this study, we build new tools and leverage volumetric imaging to examine the interplay between neuronal differentiation and the resumption of neural circuit activity during *Hydra* nervous system regeneration. We find that the ec5 neurons reach their terminal cell-fate before they functionally integrate into the *Hydra* CB circuit. These data suggest that the cues from surrounding cells direct differentiation of stem cells into the appropriate neuron subtypes rather than existing circuit activity directing these fates. Further work should be done to identify these injury-induced differentiation signals.

This work also raises the question of whether cell fate determination prior to functional recovery is true in other circuits in *Hydra*. The platform and approach developed here provide a powerful tool for answering this and other questions in *Hydra*. In particular, a major accomplishment of this work is the first demonstration of dual reporter expression in *Hydra* neurons that allows for the combined measurement of neural activity and gene expression in a live animal during regeneration. Combining this reporter system with high-speed volumetric imaging brings *Hydra* into the small but growing group of neuroscience research organisms (Lee-Liu et al., 2017; Ross et al., 2017; Zambusi and Ninkovic, 2020) in which we can perform functional volumetric imaging with cell-type specificity. To apply this method to other circuits in *Hydra*, future work should identify quantitative measures of neural activity that indicate functional recovery in each circuit. The ability to image neuronal activity throughout the animal during behavior along with cell type specific labeling will allow one to collect the types of data needed to identify normal neural circuit activity. Overall, these transgenic and imaging tools combined

with *Hydra's* unique regenerative ability differentiates it with respect from other research organisms as a model for studying complete neural circuit regeneration.

MATERIALS AND METHODS

Generation of *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$ transgenic strain

Hydra transgenic line *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$ was created by microinjecting a single plasmid containing two promoters and two transgenes. Nuclear tdTomato was driven by a 2022 bp section of the *hym176c* regulatory region, which should be specifically expressed in ec5 neurons found in the peduncle (Figure A.1 H), and GCaMP7s was driven by a 1901 bp section of the *tba1c* regulatory region, which was validated in Primack et al. (2023) to be a pan-neuronal promoter. The plasmid injection solution was injected into *Hydra vulgaris* AEP 1-cell stage embryos using an Eppendorf FemtoJet 4x and Eppendorf InjectMan NI 2 microinjector (Eppendorf; Hamburg, Germany) under a Leica M165 C stereo microscope (Leica Microscopes, Inc; Buffalo Grove, Il).

Hatchlings were screened to select tdTomato-positive polyps. Continuous asexual reproduction cycles of hatchlings with mosaic transgenic tissue yielded transgenic animals with uniform gene expression. The DNA plasmid was designed by the Robinson Lab (Rice University) and the transgenic strain was developed by Celina Juliano's Laboratory (University of California, Davis) following an established protocol (Juliano et al., 2014).

Hydra strain maintenance

All animals were maintained using standard procedures (Lenhoff and Brown, 1970). All experiments were performed using the transgenic *Hydra* line *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$. *Hydra* polyps were cultured at standard conditions, incubated at 18°C with *Hydra* media under 12hr:12hr light:dark light cycles. *Hydra* media was made with 1000X dilution of 1.0M $CaCl_2$, 0.1M $MgCl_2$, 0.03M KNO_3, 0.5M $NaHCO_3$, 0.08M $MgSO_4$. Polyps were fed three times per week with freshly hatched *Artemia nauplii* (Brine Shrimp Direct) and cleaned 6 hours post feeding with *Hydra* media. The animals were starved 24h hours prior to surgical resections.

Animal resections

Hydra were placed in a petri dish filled with *Hydra* media prior to the incision. Animal resections were performed using a scalpel, making a single incision across the center of the body, removing the aboral end and keeping the oral end to track the foot regeneration. Resectioned *Hydra* were kept at 18°C for 4 hours to allow wound closure before subjecting them to imaging procedures.

Imaging configuration

Imaging was performed by placing a resected *Hydra* between two glass coverslips separated by a 100 um spacer. Dual-color volumetric imaging was performed in 20 minute sessions at a 4 hour interval for 44 hours post amputation (hpa) at 100 fps (1.3 VPS) using Swept Confocally Aligned Planar Excitation (SCAPE) 2.0 microscopy. The system was built following the design and configuration from Voleti et al. (2019) with assistance and support from Elizabeth Hillman's

laboratory at Columbia University. The configuration consisted of an 20X Olympus (XLUMPLFLN 20XW 20x/1.00NA) as the primary objective lens, (for specimen illumination and light collection), followed by a Nikon 20x/0.75NA and Nikon 10x/0.45NA as the second and third objective lenses according to the SCAPE system nomenclature. The system used in all experiments had an effective detection NA of 0.23 and used Andor Zyla 4.2+ as the detector for imaging sessions. The microscope system provided oblique light sheet illumination across the field of view (800um × 350um × 100m). Oblique illumination was achieved by enabling the light sheet to enter the back aperture of the primary objective lens with an offset of 7 mm from the center of the objective. Coherent Obis LX 488 nm and 561 nm lasers were used as excitation laser sources for green (GCaMP7s) and red (tdTomato) channels respectively at an output power of 5mW/mm^2. Excitation and emission filters used for all dual color imaging experiments are listed in Table A.1. Epifluorescence microscopy was used for whole-animal imaging (Figure A.1).

Image processing and cell tracking

For injured *Hydra*, both channels were acquired and registered to one another and exported to 16 bit tiff format using a custom MATLAB GUI provided by Elizabeth Hillman's laboratory at Columbia University. Imaris software and 3D Viewer plugin from Fiji (Schindelin et al., 2012) were used to visualize the channel-merged image sequence in 3D. Then, maximum intensity projections from recordings were imported to Fiji and contrast was adjusted. Particle-tracking algorithm TrackMate v6.0.2 and ManualTracking (Tinevez et al., 2017) plugin from Fiji were used to track single cell activity. Prior to tdTomato expression, single cell GCaMP7s time courses were manually annotated using frame to frame analysis (ManualTracking) from Fiji.

Posterior to tdTomato expression, GCaMP7s traces corresponding to tdTomato-negative neurons continued to be manually annotated using ManualTracking while GCaMP7s traces from tdTomato-positive neurons were automatically tracked using TrackMatev6.0.2 (LoG detector sigma:15-20; Simple LAP tracker) followed by manual corrections.

For uninjured *Hydra,* both channels were acquired and registered to one another and exported to 16 bit tiff format using a custom MATLAB GUI provided by Elizabeth Hillman's laboratory at Columbia University. Imaris software was used to visualize the channel-merged image sequence in 3D. Autoregressive motion with MaxGapSize=3 was used for particle tracking. Cells that fall outside the FOV for more than 50% of the recorded time were not included in the tracking process.

Statistical analysis of neural activity

For all animals (n = 5) single cell calcium activity traces from all regeneration time points (8 - 44 hpa, Figure A.4) were analyzed using MATLAB (Figure A.3). Relative GCaMP7s intensity was normalized using the minimum and maximum pixel intensity. To evaluate the level of synchrony in the neurons of interest, all obtained traces were compared to each other by measuring the linear dependence between two arbitrary cells' activity and assigning a Pearson correlation coefficient (calculated with the MATLAB function corrcoef). The linear relationship between two arbitrary time series was performed on 15 min time series. From a scale of 0 to 1, higher value indicates higher correlation. The obtained correlation coefficients were used to build a correlation matrix for every regeneration time point. To compare the difference of correlation

coefficients within and between the regeneration time points we performed an unpaired Student's t test with bonferroni corrections.

FIGURES

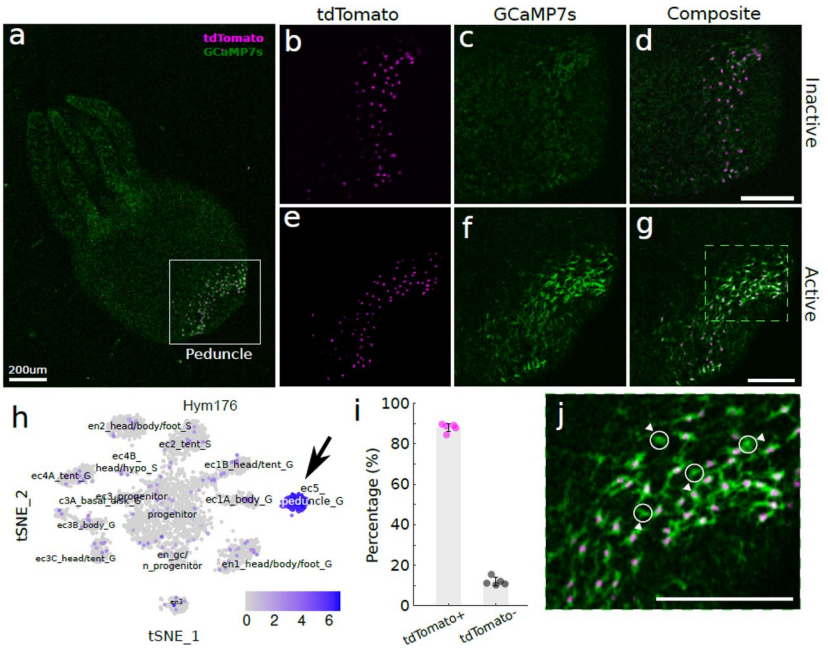


Figure A.1. ec5 neurons are part of the Contraction Burst (CB) neural circuit
(A) Fluorescence image of a *Tg(hym176c:tdTomato,tba1c:GCAMP7s)*$^{cj1\text{-}in}$*Hydra* expressing nuclear-localized tdTomato in a subpopulation of neurons in the peduncle and GCaMP7s in all the neurons. Scale bar: 200um. (B-G) High magnification images of *Hydra*'s peduncle showing dual expression of tdTomato-positive neurons in magenta and GCaMP7s in green. Scale bar: 200um. (B) Nuclear tdTomato expression in the ec5 neuronal population during inactive state. (C) GCaMP7s expression in the peduncle neurons during inactive state. (D) Composite image of (B) and (C). (E) Nuclear tdTomato expression in the ec5 neuronal population during active state. (G) GCaMP7s expression in the peduncle neurons during active state. (G) Composite image of (F) and (G). (H) t-SNE representation of single neuron transcriptomes collected from *Hydra* (Siebert et al., 2019). Arrow indicates the ec5 neurons highly expressing *hym176c*. (I) The percentage of tdTomato-positive (magenta, mean = 88.02, SD = 2.08) and tdTomato-negative (black, mean = 11.98, SD = 2.08) neurons that express GCaMP7s in the peduncle. Error bars show standard deviation. (J) High magnification of (G). Circles indicate tdTomato-negative neurons in the peduncle. Scale bar: 200um.

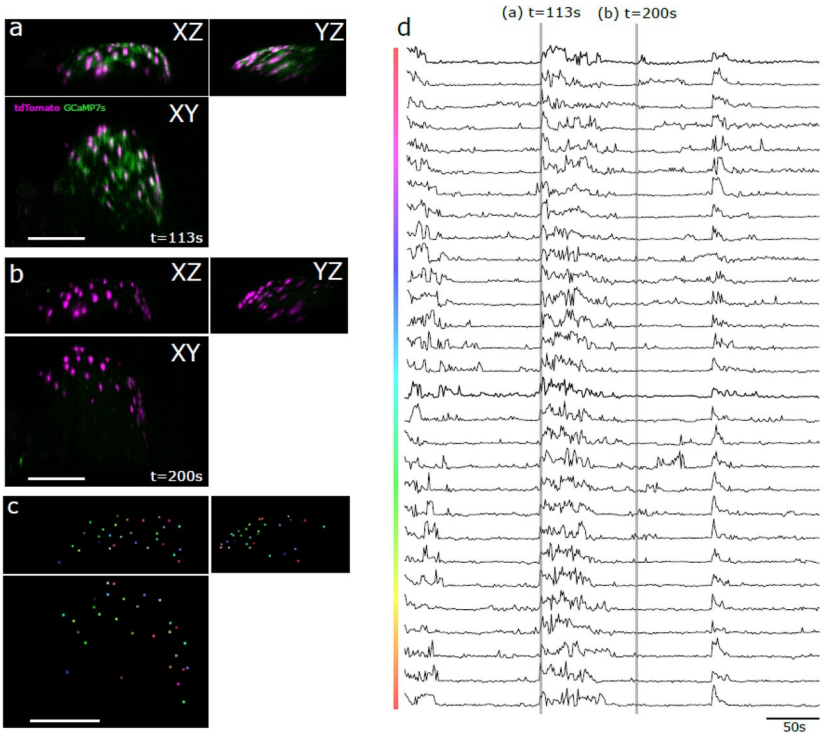


Figure A.2. SCAPE 2.0 microscopy enables volumetric tracking of cell type specific neural activity. (A-B) Maximum intensity projections along x, y, and z axes acquired from volumetric SCAPE imaging of the peduncle of a behaving *Hydra*. Each panel shows a select time during this representative imaging session. Green shows calcium activity (GCaMP7s). Magenta shows the nuclei of ec5 cells (nuclear-localized tdTomato). Scale bar = 200 um. (C) Colored dots indicate individual neurons that are tracked over the course of recording. Scale bar = 200 um. (D) Time course of the GCaMP7s fluorescence measured in each of the labeled neurons in panels a and b. The shaded regions in gray correspond to the time points used to generate the images in panels A and B.

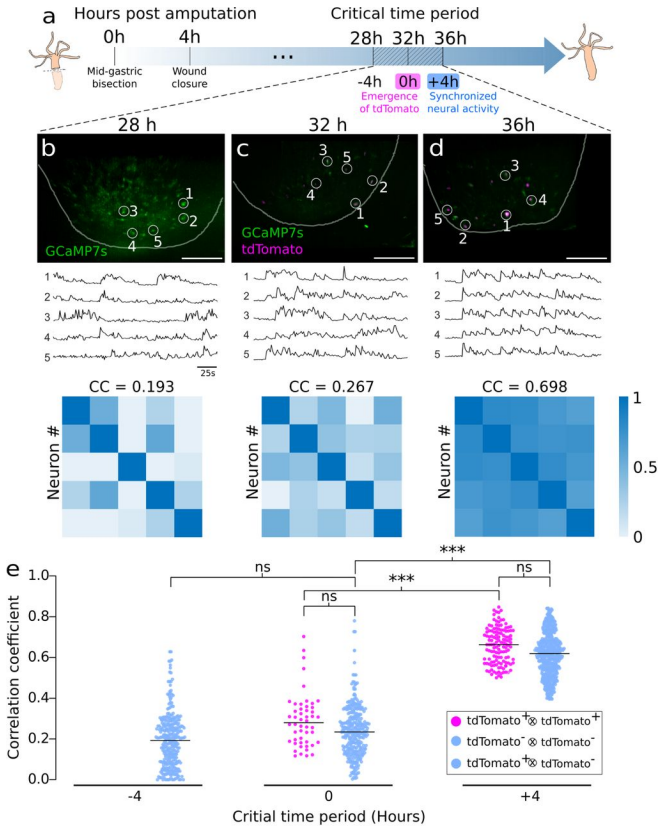


Figure A.3. ec5 neuron differentiation precedes synchronization in neural activity during peduncle regeneration. (A) Schematic representation of foot regeneration timeline after a mid-gastric bisection. (B-D) (Top row) Representative composite fluorescence image of nuclear tdTomato (magenta) and GCaMP7s (green) from the same *Hydra* at indicated time points, scale bar = 100 µm. (Middle row) GCaMP7s traces extracted from the circled neurons at time points corresponding to the top row. Numbered circles indicate the neurons of which spontaneous GCaMP7s activities are plotted. (Bottom row) Cross-correlation matrix of the circled neurons as an indicator for synchrony at time points corresponding to the top and middle row with average correlation coefficient (CC). (E) Dots represent correlation levels between two arbitrary neurons' activity. Correlation between two tdTomato-positive neurons is labeled pink. Correlation between two tdTomato-negative neurons, or between one tdTomato-positive and one tdTomato-negative neuron is labeled blue. (ns = not significant, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, Student's t test with bonferroni correction).

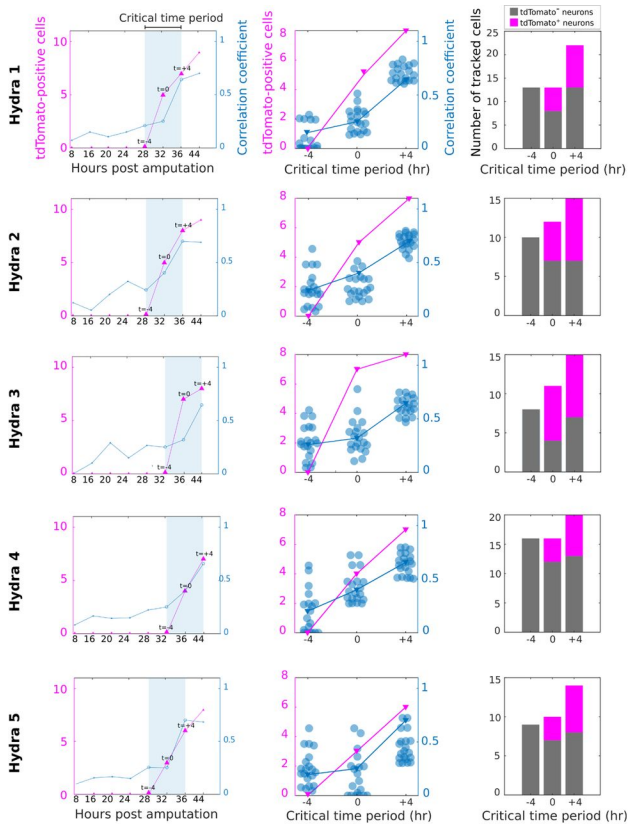


Figure A.4. Cell count and synchrony in neural activity for n = 5 *Hydra*. (Left column) The number of TdTomato-positive neurons in magenta, and the average correlation coefficient (CC) tracked over the course of 8 - 44 hpa. The blue shaded region indicates the critical time period. (Middle column) The number of tdTomato-positive neurons in magenta and the CC values in blue with average shown in inverted triangle during the critical time period. (Right panel) The number of tracked cells during the critical time period.

Supplemental tables and data can be found at DOI: 10.1101/2023.03.19.533365

Appendix B: A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state

This chapter was originally published in *Genome Research*:

Cazet JF, S Siebert, H Morris Little, P Bertemes, AS Primack, P Ladurner, M Achrainer, MT Fredrikse, RT Moreland, S Singh, S Zhang, TG Wolfsberg, CE Schnitzler, AD Baxevanis, O Simakov, B Hobmayer, CE Juliano. 2023. A chromosome-scale epigenetic map of the Hydra genome reveals conserved regulators of cell state. Genome Research *33:283-298.*

I made the following contributions to the work presented in appendix B: I collected, sequenced, and mapped three *Hydra vulgaris* AEP ATAC-seq libraries to the new genome and tested various iterations of new genome models via remapping of previously collected single cell data. I also wrote code, methods, and provided feedback on the manuscript. I specifically contributed to figures B.4, B.9, and B.10.

ABSTRACT

The epithelial and interstitial stem cells of the freshwater polyp *Hydra* are the best-characterized stem cell systems in any cnidarian, providing valuable insight into cell type evolution and the origin of stemness in animals. However, little is known about the transcriptional regulatory mechanisms that determine how these stem cells are maintained and how they give rise to their diverse differentiated progeny. To address such questions, a thorough understanding of transcriptional regulation in *Hydra* is needed. To this end, we generated extensive new resources for characterizing transcriptional regulation in *Hydra*, including new genome assemblies for *Hydra oligactis* and the AEP strain of *Hydra vulgaris*, an updated whole-animal single-cell RNA-seq atlas, and genome-wide maps of chromatin interactions, chromatin accessibility, sequence conservation, and histone modifications. These data revealed the existence of large kilobase-scale chromatin interaction domains in the *Hydra* genome that contain transcriptionally

coregulated genes. We also uncovered the transcriptomic profiles of two previously molecularly uncharacterized cell types: isorhiza-type nematocytes and somatic gonad ectoderm. Finally, we identified novel candidate regulators of cell type–specific transcription, several of which have likely been conserved at least since the divergence of *Hydra* and the jellyfish *Clytia hemisphaerica* more than 400 million years ago.

INTRODUCTION

The advent of highly specialized cell type–specific transcriptional programs played a critical role in the emergence and subsequent diversification of animal life. Decades of research have greatly advanced our understanding of the mechanisms of transcriptional regulation that underlie cell identity in metazoans. However, much of that understanding is based on findings from bilaterian species. Consequently, relatively little is known about transcriptional regulation in nonbilaterian metazoans.

Cnidaria is the sister phylum to Bilateria (Dunn et al., 2014), and despite having diverged more than 500 million years ago, the two clades show extensive homology at the molecular level. These similarities include important aspects of transcriptional regulation: Both cnidarians and bilaterians use combinatorial histone modifications and distal enhancer-like *cis*-regulatory elements (CREs) (Murad et al., 2021; Reddy et al., 2020; Schwaiger et al., 2014); many transcription factors (TFs) in bilaterians are also present in cnidarians (Babonis and Martindale, 2017; Putnam et al., 2007; Technau et al., 2005); and the target genes of developmentally significant TFs are at least partially conserved across the two clades (Gufler et al., 2018; Hartl et al., 2019; Münder et al., 2010). However, beyond these general similarities, little is known about

cnidarian gene regulatory networks and the mechanisms they use to specify and maintain cellular identity. Given Cnidaria's phylogenetic position within Metazoa, research in cnidarians is uniquely positioned to shed light on the evolutionary origins of Bilateria. In addition, many cnidarians possess remarkable abilities of self-repair and self-renewal not found in most bilaterian model systems, with species capable of whole-body regeneration (Bradshaw et al., 2015; Darling et al., 2005; Trembley et al., 1744) and potentially biological immortality (Martínez, 1998; Piraino et al., 1996; Schaible et al., 2015). Thus, a thorough characterization of transcriptional regulation in cnidarians can contribute to our understanding of both the origins and fundamental principles of transcriptional regulation of cell type in metazoans and the molecular basis for cnidarian resilience.

Species belonging to the genus *Hydra* are among the longest-studied and best-characterized cnidarian models, with the first experiments in *Hydra* dating back to 1744 (Trembley et al., 1744). *Hydra* has since been used to study patterning (Browne, 1909; Gierer and Meinhardt, 1972), stem cell biology (Bode et al., 1987; Bosch and David, 1987; David, 2012; David and Murphy, 1977), aging (Martínez, 1998; Schaible et al., 2015), regeneration (Trembley et al., 1744), and symbiosis (Fraune and Bosch, 2007; Hamada et al., 2018).

One of the strengths of *Hydra* as a research organism is its simplicity. In contrast to the three life cycle stages—planula, polyp, and medusa—found in their close cnidarian relatives, *Hydra* species possess only a polyp stage. This polyp is organized along a single oral–aboral axis, with a head made up of a mouth surrounded by a ring of tentacles at the oral pole and an adhesive foot at the aboral pole. Between the head and foot lies the body column, which

serves as both the gut and stem cell compartment. The body is made up of two epithelial layers—endoderm and ectoderm—separated by an extracellular matrix. Interspersed throughout both epithelial layers are interstitial cells, which include gland cells, neurons, germ cells, and nematocytes—the specialized stinging cells unique to cnidarians. In adult polyps, ectodermal, endodermal, and interstitial cells constitute three different cell lineages, each supported by their own stem cell population. The simplicity of this system has allowed researchers to identify every cell type in *Hydra* as well as the developmental trajectories that give rise to them (David, 2012; Siebert et al., 2019). However, the gene regulatory networks that coordinate these differentiation events remain poorly understood.

Over the past 15 years, the advent of powerful tools and resources—including a reference genome (Chapman et al., 2010), a single-cell gene expression atlas (Siebert et al., 2019), knockdown techniques (Hemmrich et al., 2012; Khalturin et al., 2008), and transgenesis (Wittlieb et al., 2006)—has allowed researchers to address topics such as regeneration and patterning at the molecular level. However, complicating the effective use of these tools is the fact that these resources were developed using different genetic backgrounds. Specifically, the currently available and recently improved reference genome (Simakov et al., 2022) was generated using strain 105 of *Hydra vulgaris* (formerly *H. magnipapillata*), whereas all transgenic *Hydra* lines and the single-cell expression atlas were generated using the AEP strain. The AEP and 105 strains belong to two distinct lineages that split ∼16 million years ago, leading to significant sequence divergence that markedly reduces cross-strain mapping efficiencies (Martínez et al., 2010; Schenkelaars et al., 2020; Schwentner and Bosch, 2015; Siebert et al., 2019; Wong et al., 2019). This highlights the need for an AEP strain reference genome that

would allow researchers to more effectively leverage transgenesis and the single-cell expression atlas.

Another appealing, although currently underused, strength of *Hydra* is that it is relatively closely related to several other established and emerging laboratory models belonging to the class Hydrozoa, creating opportunities for comparative studies. Recently published genomic and transcriptomic resources, including reference genomes for the green *Hydra viridissima* (Hamada et al., 2020) and the jellyfish *Clytia hemisphaerica* (Leclère et al., 2019) as well as a single-cell gene expression atlas of the *C. hemisphaerica* medusa (Chari et al., 2021) , provide valuable reference points for systematic comparative analyses. The *Hydra* genus is associated with several noteworthy evolutionary gains and losses, including the loss of a medusa stage, the acquisition of stably associated endosymbionts in *H. viridissima* (Schwentner and Bosch, 2015), and the loss of certain types of aboral regeneration in *Hydra oligactis* (Figure B.6; Grens et al., 1996; Hoffmeister, 1991; Weimer, 1928).Thus, effectively establishing a framework for systematic comparative approaches would greatly enhance our ability to interrogate both the conserved and unique aspects of *Hydra* biology.

To facilitate comparative genomic research in *Hydra*, we report two new high-quality genomes, a chromosome-level assembly for the AEP strain of *H. vulgaris* and a draft assembly for the *H. oligactis* Innsbruck female12 strain. To leverage these new references to better understand transcriptional regulation in *Hydra*, we used multiple independent approaches, such as assay for transposase accessible chromatin using sequencing (ATAC-seq), cleavage under targets and tagmentation (CUT&Tag) targeting histone modifications, and phylogenetic footprinting, to

annotate CREs in the AEP genome. We also generated Hi-C data that revealed domains of elevated chromatin contact frequency that likely contain transcriptionally coregulated genes. To accompany these new resources, we generated an updated and improved version of the *Hydra* single-cell atlas using the AEP-strain genome as a reference and subsequently uncovered two previously molecularly uncharacterized cell types: somatic gonad ectoderm and mature isorhiza nematocytes. We then combined our CRE annotations with the AEP single-cell atlas to identify novel candidate regulators of cell type–specific gene coexpression. Finally, we aligned the *Hydra* single-cell atlas with a *Clytia* medusa single-cell atlas and identified gene regulatory modules in the interstitial lineage that have likely been conserved over at least 400 million years of evolution (Dohrmann and Wörheide, 2017; Schwentner and Bosch, 2015). The resources generated in this study, which include a genome browser for the *H. oligactis* and strain AEP *H. vulgaris* assemblies, a BLAST server, and an interactive portal for the AEP-mapped *Hydra* single-cell atlas, are available at the *Hydra* AEP Genome Project Portal (*https://research.nhgri.nih.gov/HydraAEP/*).

RESULTS

Generation and annotation of two high-quality *Hydra* genome assemblies

We sequenced, assembled, and annotated a chromosome-level genome assembly for the AEP laboratory strain of *H. vulgaris* (for details, see Supplemental Material; Figures B.6 and B.7; Table B.1; Supplemental Data S1). In addition, we generated a high-quality draft genome for the Innsbruck female12 strain of *H. oligactis*. We were motivated to generate a genome reference for *H. oligactis* because its phylogenetic position as a sister species to *H. vulgaris*—along with unique traits such as reduced regenerative capacity (Grens et al., 1996; Hoffmeister, 1991;

Weimer, 1928), a deficient heat shock response (Bosch et al., 1988), and inducible senescence (Yoshida et al., 2006)—makes it valuable for comparative genomic studies of the *Hydra* genus. The resulting assemblies for *H. vulgaris* and *H. oligactis* were of equivalent or greater completeness and contiguity compared with other available hydrozoan genomes (Figure B.1 A; Table B.2).

We found that synteny in the strain 105 and AEP genome assemblies was highly conserved, with the notable exception of an ~5-Mb inversion on Chromosome 8 (Figure B.8 A,B; Table B.3). Similarly, the centromeric repeats in the two strains were highly similar, although not identical (Figure B.8 C; Melters et al., 2013). In addition, this analysis allowed us to place nearly all (36/39) of the unincorporated scaffolds from the strain 105 assembly onto one of the 15 pseudochromosome scaffolds in the AEP assembly (Supplemental Data S2). Similarly, we were able to generate preliminary chromosome assignments for contigs covering 91.3% of the sequence (1.16 out of 1.27 Gb) in the *H. oligactis* assembly and 32.3% (91.8 out of 284.3 Mb) of a previously published assembly for *H. viridissima* (Figure B.8 D,E; Supplemental Data S2; Hamada et al., 2020).

To augment the strain AEP *H. vulgaris* genome assembly, we also generated genome-wide CRE annotations. To do this, we used the ATAC-seq (Buenrostro et al., 2013; Corces et al., 2017) to map accessible regions of chromatin. We also established a protocol for performing CUT&Tag (Kaya-Okur et al., 2019) in *Hydra* to globally map multiple histone modifications, including the repressive histone modification H3K27me3 as well as the activating histone modifications H3K4me1 and H3K4me3 (Supplemental Data S3; for details, see Supplemental Material). We

validated our results by confirming that they matched the expected distribution patterns of their associated genomic features (Figure B.1 B,C; Figures B.9-B.11).

To supplement our CRE annotations, we performed phylogenetic footprinting (Gumucio et al., 1992; Tagle et al., 1988) by using previously published genomes for the hydrozoans *C. hemisphaerica* (Leclère et al., 2019), *H. viridissima* (Hamada et al., 2020), and the 105 strain of *H. vulgaris* (Chapman et al., 2010)—along with our newly assembled genomes—to generate a cross-species whole-genome alignment that spanned ~400 million years of hydrozoan evolution (Figure B.1 B). Our alignment yielded results that recapitulated the findings from previous manual cross-species alignments of individual *Hydra* promoter regions (Figure B.12; Vogg et al., 2019), supporting the accuracy of our genome-wide approach. We then used our whole-genome alignment to classify genomic features as either conserved or nonconserved (for details, see Supplemental Material). We provide lists of conserved noncoding genomic features in Supplemental Data S3.

Prediction of conserved TF binding sites using phylogenetic footprinting

Accurately identifying TF binding sites in CREs is an essential, albeit often challenging, aspect of gene regulatory network characterization. This task is made especially difficult in nonbilaterian metazoans by the lack of specific antibodies needed for conventional TF mapping assays (e.g., ChIP-seq and CUT&RUN). The lack of binding data can even hinder computational approaches for predicting binding sites, as the binding preferences of cnidarian TFs typically must be inferred from data collected from distantly related bilaterians. We therefore sought to evaluate the functional relevance of bilaterian TF binding motifs in *Hydra* by leveraging

phylogenetic footprinting to determine which motifs showed evidence of conservation. Of the 840 motifs considered in our analysis, we found that 384 (45.7%), including those that are bound by numerous conserved and developmentally significant TFs (Figure B.2 A), had significantly higher genome-wide conservation rates compared with the shuffled controls (Supplemental Data S4). This suggests that there is extensive conservation of TF binding preferences from cnidarians to bilaterians.

Another confounding issue for ab initio TF binding site predictions is that TF binding motifs are typically short and degenerate, leading to high false-positive rates. However, by filtering putative TF binding sites using both our ATAC-seq and phylogenetic footprinting data, we reduced the total number of predicted binding sites genome-wide by >99%, from more than 45 million to 210,122 (Supplemental Data S5). Thus, we simplified the landscape of putative TF binding sites by eliminating loci with a relatively low probability of being bona fide binding sites.

Many *Hydra* genes are likely regulated by distal regulatory elements

In bilaterians, transcriptional regulation frequently involves long-range interactions between distal CREs and their target promoter, often spanning dozens of kilobases. However, numerous successful reporter lines have been generated in *Hydra* using only 500–2000 bp of flanking sequence upstream of a gene of interest, motivating some to hypothesize that transcriptional regulation in *Hydra* is simpler than in bilaterians and primarily regulated by promoter-proximal elements that typically fall within 2 kb of the TSS (Klimovich et al., 2019). However, this hypothesis has not been systematically investigated.

To better understand the distribution of CREs in the *Hydra* genome, we used our cross-species whole-genome alignment to characterize sequence conservation rates around genes in the AEP assembly (Figure B.2 B). We found that flanking noncoding sequences around genes had elevated conservation rates that extended ~4.4 kb upstream and ~2.8 kb downstream before falling back to baseline levels. Although we found that most of the elevated sequence conservation fell within 2 kb upstream of the TSS, nearly half of the conservation signal fell outside of that boundary (Figure B.2 B). In addition, we found that ~44% of genes in our analysis had at least one conserved ATAC-seq or H3K4me1 peak further than 2 kb upstream of the TSS (Figure B.2 C,D). These results indicate that there are likely many instances in which functionally important CREs lie further than 2 kb from their target gene and highlight the need for functional genomic data to accurately identify promoter regions.

Hydra chromatin is organized into localized contact domains

The three-dimensional organization of DNA molecules in the nucleus is tightly linked to genome regulation (Szabo et al., 2019). Although several cnidarian Hi-C data sets have been published (Supplemental Table S4; Zimmermann et al., 2022; Li et al., 2020; Nong et al., 2020; Simakov et al., 2022) ,the 3D organization of cnidarian genomes remains largely uncharacterized. We therefore interrogated our *Hydra* Hi-C data to better understand the 3D architecture of the *Hydra* genome.

We first examined chromatin interactions at the whole-chromosome scale. We observed signatures of a Rabl-like conformation (Hoencamp et al., 2021), with interactions occurring between centromeres of different chromosomes as well as between centromeres and telomeres

within individual chromosomes (Figure B.3 A). Compared with previously characterized cnidarian genomes, these interaction patterns appeared unique to *Hydra*, as we had not observed similar phenomena in other publicly available cnidarian Hi-C data sets. We therefore performed a systematic analysis of inter-chromosomal interactions in cnidarians (for details, see Supplemental Materials) and found that the *Hydra* genome had significantly elevated levels of inter-centromeric interactions, but not inter-telomeric interactions, relative to other cnidarians (Figure B.13). Notably, this change in 3D genome organization appeared to be correlated with the loss of multiple condensin II subunits in hydrozoans (Figure B.14; Supplemental Data S6). These lost subunits were shown to inhibit inter-chromosomal interactions in other species (Hoencamp et al., 2021), suggesting that their loss has resulted in the elevated levels of inter-centromeric interactions in *Hydra* and possibly other hydrozoans. However, the extent to which these interaction patterns are present in other hydrozoan genomes is unknown owing to a lack of Hi-C data from other hydrozoan species.

We next explored intra-chromosomal interactions in our Hi-C data to look for evidence of chromatin domains or loops, which are structures generated by transcriptional regulatory mechanisms in diverse eukaryotic genomes (Szabo et al., 2019; Zheng and Xie, 2019). We found that *Hydra* chromatin is hierarchically organized into megabase-scale domains that contain much smaller kilobase-scale subdomains (Figure B.3 B-D). The larger megabase-scale domains showed a checkerboard-like interaction pattern consistent with the A/B compartments observed in other Hi-C data sets (Figure B.3 B; (Zheng and Xie, 2019). Within these A/B compartments were more localized structures that in places resembled the triangle-shaped patterns associated with topologically associating domains (TADs) in other species (Figure B.3 C,D). We also

occasionally observed contact patterns suggestive of chromatin loops (Figure B.15), but such structures were rare.

To determine if the contact domains we observed were associated with transcriptional regulation, we first used a previously established computational pipeline (Ramírez et al., 2018) to predict chromatin domain boundaries (Figure B.3 D). Although the resolution of our Hi-C data made it difficult to fully resolve the kilobase scale domains apparent in the *Hydra* genome, we were nonetheless able to identify 4028 putative contact domains across the AEP assembly with a median size of ~176 kb using this approach (Supplemental Data S7). We then used the *Hydra* single-cell atlas (described below) to characterize the expression patterns of genes around the predicted domain boundaries. We found that the cell type–specific expression patterns of adjacent gene pairs that fell within the same contact domain were significantly more correlated than adjacent gene pairs that spanned a domain boundary (Figure B.3 E), suggesting that *Hydra* chromatin contact domains are indeed associated with transcriptional regulation. We also found that chromatin boundaries were depleted of several euchromatin markers—including chromatin accessibility, sequence conservation, and H3K4me1—and enriched in heterochromatin markers such as increased repetitive element density and higher levels of H3K27me3 (Figure B.3 F-J). Altogether, these results suggest that similar to invertebrate bilaterians, *Hydra* chromatin is organized into large epigenetically regulated domains that contain coregulated genes. In addition, the clear correlation between the predicted location of domain boundaries and other orthogonal data sets such as ATAC-seq and CUT&Tag shows that our domain prediction analysis indeed captured meaningful aspects of chromatin architecture across the *Hydra* genome.

An updated single-cell RNA-seq atlas for *H. vulgaris* uncovers the transcriptional profiles of additional cell types

We next used the genomic resources we had generated to interrogate the transcriptional regulation of cell type specification in *Hydra*, which required access not only to CRE annotations but also to the transcriptomic profiles associated with different *Hydra* cell types. We previously published a whole-animal single-cell RNA-seq (scRNA-seq) data set for the AEP strain of *H. vulgaris* that provides an atlas of molecular cell states in adult polyps (Siebert et al. 2019). However, the currently available versions of this data set use either the strain 105 genome or an AEP strain transcriptome as a reference. Both are suboptimal as the transcriptome does not provide information about genomic context, thus hindering any research into transcriptional regulation, and the 105 genome gene models are less complete and have reduced mapping rates when using AEP RNA-seq data (Figure B.16; Table B.2). In addition, there have been substantial improvements in normalization (Hafemeister and Satija, 2019), batch-correction (Stuart et al., 2019), and visualization techniques (McInnes et al., 2018) for scRNA-seq data since the *Hydra* single-cell atlas was initially published. Therefore, we addressed these limitations by reanalyzing the data using the AEP assembly as a reference.

Following mapping and doublet removal (for details, see Supplemental Materials; Figures B.17 and B.18), we recovered 29,339 single-cell transcriptomes that passed our quality control cutoffs, an increase of ~17.4% compared with the 24,985 transcriptomes presented in the originally published atlas (Siebert et al., 2019). We then used Seurat to perform a Louvain clustering analysis and visualized the results using a uniform manifold approximation and projection

(UMAP) dimensional reduction (Figure B.4 A; (Hao et al., 2021a; McInnes et al., 2018; Waltman and Van Eck, 2013). We then annotated the resulting clusters using established cell type markers (Figure B.19; Siebert et al., 2019). While generating these annotations, we identified two cell types that were not found in previous iterations of the single-cell atlas: isorhiza-type nematocytes and ectodermal male and female somatic gonad cells. We subsequently identified markers of these two populations, which we validated using in situ hybridization (Figure B.4 B-H). The isorhiza marker, *G008733*, has no known functional domains and appears to be specific to brown *Hydra*. The somatic gonad marker, *parascleraxis* (*G017021*), is the ancestral ortholog of two paralogous vertebrate basic helix-loop-helix TFs, *paraxis*/*tcf15* and *scleraxis*, that regulate muscle differentiation (Della Gaspera et al., 2022; Freitas et al., 2006).

In summary, we generated an updated scRNA-seq atlas for whole adult *Hydra* that can now be used in conjunction with the AEP genome assembly. This comprehensively annotated atlas, which incorporates two additional cell types, contains virtually all known cell types in an adult *Hydra*. We also provide exhaustive lists of marker genes for all clusters (Supplemental Data S8) as well as 56 modules of co-expressed genes (Figure B.20; Supplemental Data S9, S10).

Characterizing the evolutionary history of *Hydra* cell type–specific transcriptomes

The *Hydra* single-cell atlas captures the transcriptional signatures of virtually all cell states in an adult polyp, which presents a valuable opportunity to gain new insight into the evolutionary history of the transcriptional programs that define cnidarian cell types. The acquisition of novel

cellular traits is often accompanied by a concurrent period of genetic innovation (Arendt, 2008; Khalturin et al., 2009). This can leave a phylogenetic signature in a cell's transcriptome in the form of an overrepresentation of novel genes that arose during periods of evolutionary change in a cell type's transcriptional program (Domazet-Lošo et al., 2007). Thus, characterizing the age distribution of genes expressed in different cell types can shed light on when those genetic programs arose.

To analyze the relationship between gene age and transcriptional specificity, we first assigned phylostratigraphic ages to *Hydra* gene families using orthology predictions generated from an OrthoFinder analysis of 44 metazoan proteomes (Figure B.21; Supplemental Table S5; Supplemental Data S11; Emms and Kelly, 2019, 2015). We then characterized the relative enrichment of genes of a given age across different cell types in our scRNA-seq atlas, revealing clear cell type–specific enrichment patterns (Figure B.22 A). We also calculated a holistic score, the transcriptome age index (TAI) (Domazet-Lošo and Tautz, 2010), for each cell cluster (Figure B.22 B,C). Consistent with previous reports (Hemmrich et al., 2012), we found that ancient gene families predating Metazoa were most strongly associated with interstitial cells that have a high degree of potency, namely, interstitial stem cells, early neuron and nematocyte progenitors, and germ cells—with interstitial stem cells having the least derived transcriptomic profile overall (Figure B.22).

Among differentiated interstitial cell types, both gland cells and neurons were enriched for genes that originated at the base of Metazoa, likely reflecting the ancient origins of their respective transcriptional programs (Figure B.22; Musser et al., 2021; Smith and Mayorova, 2019).

However, neurons also showed enrichment for younger genes, suggesting the existence of cnidarian-specific modifications to neuronal transcription. In contrast, nematocyte transcriptional profiles were generally younger, with nematoblasts (i.e., developing nematocytes) showing stark enrichment for gene families that originated either at the base of Cnidaria or Medusozoa (Figure B.22), consistent with the more recent evolutionary origin of nematocytes (David et al., 2008; Hwang et al., 2007). The two epithelial lineages were both associated with genes predating Cnidaria, although endodermal cell transcriptomes appeared somewhat older than those in ectodermal cells (Figure B.22). Like neurons, both epithelial lineages were also enriched for younger hydrozoan-specific gene families. Overall, our analysis suggests that the transcriptional programs used by interstitial stem cells, germ cells, nematoblasts, and gland cells show relatively little genetic innovation since their initial emergence, whereas epithelial and neuronal transcriptional programs have been more dynamic over the course of cnidarian evolution.

Prediction of *Hydra* cell fate regulators

We next sought to leverage both the scRNA-seq atlas and the AEP assembly CRE annotations to identify TFs involved in coordinating *Hydra* cell type–specific transcriptional programs. We had previously explored this question as part of the initial publication of the *Hydra* atlas using an analysis that combined ATAC-seq from strain 105 polyps with the strain AEP scRNA-seq data (Siebert et al., 2019). Broadly, our approach was first to identify TF binding motifs that were enriched in promoter-proximal CREs associated with a set of co-expressed genes, collectively referred to as a metagene. Then, we predicted candidate regulators by identifying TFs that both had similar expression to the metagene of interest and could plausibly bind one of the enriched motifs. We were motivated to revisit this analysis for two reasons: First, we could use our

improved AEP-mapped atlas, and second, our phylogenetic footprinting data would improve our enrichment analysis by eliminating potential TF binding sites that were likely not functionally relevant.

Our motif enrichment analysis identified 336 motifs that were enriched in at least one metagene in the AEP-mapped *Hydra* single-cell atlas (Figures B.20, B.23; Supplemental Data S12), and our subsequent co-expression analysis identified 115 TFs as candidate regulators (Figure B.4 J-N; Figure B.22; Supplemental Data S13). These candidates spanned diverse cell states and included multiple regulators whose function had been previously validated in *Hydra*, such as TCF/Wnt signaling as a regulator of oral tissue (Figure B.4 I; Broun et al., 2005; Gee et al., 2010; Hobmayer et al., 2000; Lengfeld et al., 2009), *gata1-3* as a regulator of aboral tissue (Figure B.4 J; Ferenc et al., 2021), and *zic4* as a regulator of epithelial tentacle tissue (Figure B.24 J; Vogg et al., 2022). These results validate our analysis as a method for detecting functionally meaningful regulatory relationships underlying cell fate decisions in *Hydra*. In addition, our analysis identified novel candidate regulators. These included *pou4* as a regulator of late stage nematogenesis and neurogenesis (Figure B.4 K), *ebf* as a regulator of oogenesis (Figure B.4 L), and *nr2f-like* as a regulator of early nematogenesis (Figure B.4 M).

Systematic comparison of cell type–specific transcription in *H. vulgaris* and *C. hemisphaerica*

We next extended our analysis of cell type–specific transcription to another hydrozoan, the jellyfish *C. hemisphaerica*. *Hydra* and *Clytia*, although both hydrozoans, nonetheless show extensive differences at both the genomic and phenotypic level. The most recent common

ancestor of *Hydra* and *Clytia* lived more than 400 million years ago (Dohrmann and Wörheide, 2017; Schwentner and Bosch, 2015), and the protein sequence divergence between the two species is roughly equivalent to that of humans and lampreys (Figure B.21). *Hydra* and *Clytia* also have markedly different life cycles: *Hydra* have a derived and simplified life cycle that consists only of a polyp stage, whereas *Clytia* have planula, colonial polyp, and medusa stages, each with distinct morphologies. Because of the extensive divergence between these lineages, identifying molecular commonalities between these two systems provides strong evidence of conservation and, by extension, functional significance.

To identify conserved cell type–specific transcriptional patterns in *Hydra* and *Clytia*, we used reciprocal principal component analysis to align our *Hydra* scRNA-seq atlas to a recently published scRNA-seq atlas of the *Clytia* medusa (Chari et al., 2021). The resulting UMAP representation accurately grouped homologous cell types from the two species (Figure B.5 A-C). To assess transcriptional similarities between cell types more quantitatively, we calculated an alignment score (Tarashansky et al., 2021) for all pairwise cross-species cell type comparisons. This revealed extensive similarities between the two species, providing strong evidence of transcriptional conservation across homologous cell types (Figure. 5D). We also calculated a distance metric that quantified the overall degree of transcriptional equivalence between a given cell and similar cells in the other species (Figure B.25).

Among the three lineages, epithelial cells showed fewer cell type similarities than did interstitial cells (Figure B.5 D), consistent with the marked differences in epithelial morphology between polyp and medusa body plans. Nonetheless, we did identify some transcriptional similarities

among epithelial cells, suggesting that hydrozoan medusa and polyp body plans are created at least in part through the redeployment of shared transcriptional programs. In addition, *Hydra* epithelial stem cells had low transcriptional distance scores (Figure B.25), potentially indicating the conservation of general epithelial transcriptional signatures despite the lack of direct homologies with individual *Clytia* epithelial cell types. Interstitial cell types showed more robust conservation, with nearly all *Hydra* interstitial cell populations showing similarity to at least one *Clytia* cell type (Figure B.5 D). In some cases, there was clear one-to-one homology, such as female germline cells and some gland cell subtypes. In contrast, neuron and nematocyte cell types had either one-to-many or many-to-many patterns of homology.

The *Hydra* genus has undergone extensive gene loss, likely as a consequence of its simplified life cycle (Chapman et al., 2010; Hamada et al., 2020; Leclère et al., 2019), but the ancestral function of these lost genes has gone largely unexplored. We sought to leverage the *Clytia* cell atlas to systematically characterize the potential function of genes lost in *Hydra*. To do this, we calculated a holistic score for each *Clytia* cell cluster that represented the degree to which that cell type expressed these lost genes (Figure B.26). The *Clytia* cell type with the highest score was the tentacle GFP cell, a bioluminescent cell type located in the medusa tentacle bulb (Fourrage et al., 2014). Among other cell types, gland cell scores were clear outliers, with exceptionally high values across all subtypes. Notably, our cross-species cell type comparison found that the tentacle GFP cell, along with three of the five *Clytia* gland cell subtypes, did not show strong homology with any *Hydra* cell types (Figure B.5 D). These observations suggest that gene loss in the *Hydra* genus has been driven, at least in part, by the loss or simplification of cell type–specific transcriptional programs.

Interstitial cell–specific gene regulatory modules are conserved between *Hydra* and *Clytia*

Transcriptional similarities between *Hydra* and *Clytia* cell types imply the existence of conserved gene regulatory networks. Therefore, we sought to identify the regulators underlying conserved cell type–specific transcription in these two species. To that end, we reapplied the approach we used to identify candidate gene module regulators in *Hydra* to the *Clytia* single-cell atlas, albeit with some modifications because of the lack of epigenetic data in *Clytia* (Figure B.27; Supplemental Data S14, S15). We then compared the results from each species to identify commonalities. We found 13 motifs that had similar enrichment patterns in the two species (enrichment correlation score > 0.5) (Supplemental Data S16). Thus, despite the high level of divergence in noncoding sequence between the *Clytia* and *Hydra* genomes, we see significant overlap in the motifs associated with conserved gene co-expression modules.

To find candidate regulators of conserved gene co-expression modules, we first sought to identify TFs with similar cell type specificity in *Hydra* and *Clytia*. To do this, we identified one-to-one ortholog pairs with correlated expression in the aligned cross-species principal component space (for details, see Supplemental Material). This approach recovered 409 orthologs with highly conserved expression patterns (correlation score > 0.65), including markers for most cell types in the cross-species atlas (Figures B.28 and B.29; Supplemental Data S17). From these 409 orthologs, we identified 30 predicted TFs with conserved cell type–specific expression (Figure B.30). Although our analysis did not recover any conserved TFs in epithelial cells (likely because of the relatively poor alignability of the epithelial cell clusters), we did find putative conserved regulators for all interstitial cell types.

To further test if the function of these 30 TFs was conserved from *Clytia* to *Hydra*, we manually cross-referenced their expression patterns with our cross-species motif enrichment analysis to identify cases in which both the TF expression pattern and its binding motif enrichment profile were conserved. We identified five TFs that met these stringent conservation criteria (Figure B.5 E-H; Figure B.31), including regulators of neurogenesis (*pou4* and *atoh8*) (Figure B.5 E; Figure B.31), nematogenesis (*pou4*, *paxA*, and *foxn1/4*) (Figure B.5 E-G), and oogenesis (*ebf*) (Figure B.5 H). Thus, by systematically comparing genomic and transcriptomic data from distantly related hydrozoan species, we were able to identify transcriptional regulators of multiple interstitial cell types that have likely retained their function over at least 400 million years of evolution.

DISCUSSION

Characterizing transcriptional regulation in nonbilaterian metazoans presents significant challenges. In this study, we generated new genomes for *H. oligactis* and strain AEP *H. vulgaris*—with the latter being among the most contiguous and best-annotated cnidarian genomes currently available—to facilitate the investigation of hydrozoan transcriptional regulation. By combining our AEP strain assembly with data covering single-cell expression, chromatin accessibility, histone modifications, sequence conservation, and chromatin contact frequency, we were able to perform the most in-depth characterization of transcriptional regulation in a cnidarian to date. These new resources, available at *https://research.nhgri.nih.gov/HydraAEP/*, provide powerful new tools for future research aimed at unraveling hydrozoan transcriptional regulation.

Evidence of long-range chromatin interactions in *Hydra*

Consistent with previous characterizations of CREs in cnidarians (Murad et al., 2021; Reddy et al., 2020; Schwaiger et al., 2014), our global maps of histone modifications and chromatin accessibility clearly support the existence of distal enhancer-like regulatory elements in *Hydra*. This is further supported by our phylogenetic footprinting analysis, which found that many of these putative distal elements were conserved across multiple *Hydra* species. Nonetheless, we found that *Hydra* CREs show a strong promoter-proximal bias, with most conserved upstream elements falling within 2 kb of the TSS. This likely explains the relatively high success rate of transgenic *Hydra* reporter lines generated using just 1–2 kb of upstream sequence (Klimovich et al., 2019). However, there have been some instances, such as with *hym-176e* and β-*catenin*, in which short stretches of upstream promoter proximal sequence were not sufficient to fully recapitulate known expression patterns (Hobmayer et al., 2000; Iachetta et al., 2018; Noro et al., 2019). Therefore, the genomic resources generated by this study should facilitate the generation of transgenic reporter lines in the future by allowing researchers to identify likely promoter regions using data collected from the same strain used for transgenesis.

Our characterization of the 3D chromatin architecture of the strain AEP genome provided further evidence that distal chromatin interactions are likely prevalent in *Hydra*, as we identified thousands of localized chromatin interaction domains that spanned dozens to hundreds of kilobases. The borders of these domains were marked by changes in histone modifications and gene expression patterns, indicating that they were likely related to transcriptional regulation. Thus, *Hydra* chromatin domains resemble those found in other organisms that lack CTCF-mediated chromatin loops, such as *Drosophila* and *Arabidopsis*, where TADs arise passively via

the partitioning of heterochromatin and euchromatin into distinct interaction compartments (Rowley et al., 2017; Szabo et al., 2019). However, many of the proteins that localize to TAD boundaries in *Drosophila*, the system in which non-CTCF-mediated chromatin organization has been best characterized (e.g., BEAF-32, CP190, Chromator, GAF, and M1BP) (Szabo et al., 2019), appear to be absent from the *Hydra* genome. It therefore remains unclear how domain boundaries are regulated in cnidarians.

A highly conserved feature of chromatin domains in many organisms is that their boundaries often overlap with regions of active chromatin (Szabo et al., 2019). In stark contrast, we found that *Hydra* domain boundaries generally fell within stretches of heterochromatin. In *Drosophila*, it was proposed that active regions found at putative domain boundaries are not boundaries at all but rather are small active domains interspersed between larger repressed domains (Rowley et al., 2017). Thus, it may be the case that the heterochromatic signature found at *Hydra* domain boundaries corresponds to small, repressed regions that we are unable to resolve with our current whole-animal Hi-C data. In the future, the generation of higher resolution Hi-C data from a more homogenous cell population would help clarify the nature and regulation of *Hydra* domain boundaries.

We also used our Hi-C data to characterize the 3D organization of the *Hydra* genome at the chromosomal level, which revealed high levels of inter-centromeric interactions. Indeed, we performed a systematic cross-species analysis of available cnidarian Hi-C data sets and found that *Hydra* had significantly elevated levels of inter-centromeric chromatin interactions relative to other cnidarians, which may have resulted from the loss of a subset of condensin II subunits in

the hydrozoan lineage. Alternatively, the increased inter-chromosomal interactions may simply be a byproduct of the increased size of brown *Hydra* genomes (Wong et al., 2019). Characterizing inter-chromosomal contacts in other hydrozoans, particularly in the green *H. viridissima*, which has a much smaller genome than *H. vulgaris*, as well as *H. oligactis*, which has a larger genome, would help address this question. In addition, the Hi-C data generated from these experiments could be used to generate chromosome-level scaffolds from the available draft genomes for these two species (Hamada et al., 2020), which would greatly facilitate future comparative genomics research within the *Hydra* lineage.

Deep conservation of hydrozoan cell type–specific transcriptional programs

The stem cell differentiation trajectories in *Hydra* are the best characterized of any cnidarian, making it well suited for exploring the gene regulatory networks underlying cell fate specification. To that end, we combined the CRE annotations we generated for the AEP assembly with an updated version of the *Hydra* scRNA-seq atlas to better understand the transcriptional programs directing cell type–specific transcription. In the process of updating the atlas, we recovered ∼17% more single-cell transcriptomes and two additional cell types compared with previous atlas iterations. With the addition of these two previously absent cell types, isorhiza nematocytes and somatic gonad ectoderm, the *Hydra* single-cell atlas now contains virtually all known cell types in the adult polyp. However, there is likely additional complexity within these two additional cell populations that we are currently unable to resolve. Specifically, we currently cannot differentiate between the two types of mature isorhiza nematocytes, holotrichous and atrichous, nor can we distinguish between male and female somatic gonad. The inability to resolve these subtypes likely results from their relatively low

abundance in our data set. The generation of transgenic reporter lines using the markers we provide in this study would greatly facilitate efforts to selectively isolate and transcriptionally profile these cell subtypes.

Our subsequent analysis of the updated atlas provided several insights into the evolution of cell type–specific transcriptional programs in hydrozoans. Consistent with previously published findings (Hemmrich et al., 2012), our phylostratigraphic analysis of the three adult stem cell populations found that the genes transcribed in the epithelial stem cells are substantially younger than those transcribed in interstitial stem cells. Indeed, the transcriptional profiles of the two epithelial stem cell populations and their differentiated progeny were enriched in genes originating at the base of hydrozoa or later. Little is known about the evolution of epithelial cells within hydrozoans, but the topic may merit further study as our analysis suggests these cell types may be a major driver of recent genetic novelty. In contrast, interstitial stem cells had the oldest transcriptional profile of any cell in the *Hydra* atlas. This finding is consistent with the proposed existence of a deeply conserved genetic program underlying pluripotency in metazoans (Juliano et al., 2010; Sogabe et al., 2019). Thus, although interstitial stem cells are thought to be a derived cell type (Gold and Jacobs, 2013), they make use of an evolutionarily ancient transcriptional program.

To better understand the regulation of these cell type–specific transcriptional programs, we integrated our CRE annotations and scRNA-seq data to identify candidate TFs involved in cell fate specification. This analysis recovered previously characterized as well as novel candidate regulators, thus providing an extensive list of candidates for future functional studies across a

diverse array of cell types. Notably, in addition to capturing the known functions of previously characterized TFs, our analysis also predicted additional functions that may have been missed by previous studies. Specifically, our work suggests that *gata1-3* and *zic4* regulate transcription in neurons (Figure B.4 J; Figure B.24 J) in addition to their previously documented roles in epithelial cells (Ferenc et al., 2021; Vogg et al., 2022).

To determine if the composition and regulation of cell type–specific transcription in *Hydra* are conserved in other hydrozoans, we performed a systematic comparative analysis of the *Hydra* and *Clytia* single-cell atlases. This analysis revealed extensive conservation in cell type–specific transcriptional signatures despite the extensive divergence between these two species, which allowed us to identify hundreds of conserved marker genes across all major cell types. However, apart from germ cells and a subset of gland cells, we did not observe clear one-to-one homology among differentiated cell subtypes. This may indicate that although broad cell types (e.g., neuron, nematocyte, gland cell) are well conserved at the transcriptional level, the identities of specific subtypes are not. Indeed, some *Clytia* cell types, including the tentacle GFP cell and several gland cell subtypes, appear to have been lost in *Hydra*. One possible hypothesis is that this loss was driven by the simplification of the *Hydra* life cycle. However, it is currently unclear if such a hypothesis is plausible, as it is not known if the cell types in question are medusa-specific in *Clytia*. The generation of single-cell atlases for the *Clytia* planula and polyp stages would help address this question.

Among the cell types that were clearly conserved between *Hydra* and *Clytia*, our analysis uncovered robust overlap not only in gene expression but also in predicted transcriptional

regulators. Specifically, we identified putative regulators of nematogenesis, neurogenesis, and oogenesis whose gene expression patterns and motif enrichment profiles were conserved from *Clytia* to *Hydra*. Given the extensive transcriptional similarities we observed in our aligned cross-species atlas, it is very likely that the relatively small list of conserved regulators we identified in this study is incomplete. The generation of CRE annotations for the *Clytia* genome would likely increase the sensitivity of this analysis and help reveal additional regulatory conservation.

Among the TFs we identified as having a conserved function in hydrozoans, three of these regulators—namely, *pou4*, *atoh8*, and *paxA*—have been functionally characterized in *Nematostella*, a cnidarian that diverged from hydrozoans more than 600 million years ago (Dohrmann and Wörheide, 2017; Schwentner and Bosch, 2015). In all three cases, the reported roles of these TFs in *Nematostella* are consistent with their predicted functions in hydrozoa based on our analysis (Babonis and Martindale, 2017; Richards and Rentzsch, 2015; Tournière et al., 2020). In addition, our predictions regarding *pou4* and *atoh8*function in hydrozoan neurons are consistent with the well-established roles for these genes in bilaterian nervous systems (Gan et al., 1996; Inoue et al., 2001). Collectively, these findings support the accuracy of our analytical approach and provide insight into the likely ancestral function of these TFs in the last common cnidarian ancestor.

Our analysis also identified novel regulators, including *foxn1/4* as a regulator of nematocyte maturation and *ebf* as a regulator of oogenesis. Although putative functions for these TFs have not, to our knowledge, been previously described, we did find publicly available expression data

sets that were consistent with *ebf* having a conserved role in oogenesis. Specifically, recently published bulk RNA-seq data from the hydrozoan *Hydractinia symbiolongicarpus* showed that *ebf* was specifically expressed in polyps undergoing oogenesis (DuBuc et al., 2020). In addition, an scRNA-seq atlas of the zebrafish ovary shows an *ebf* ortholog, *ebf3b*, as a marker of female germline stem cells (Liu et al., 2022). Therefore, *ebf* regulation of oogenesis may predate the split of Bilateria and Cnidaria.

In summary, by taking a comparative approach and leveraging the genomic and transcriptomic data available in *Clytia* and *Hydra*, we identified both conserved gene co-expression modules and the TFs that likely regulate them, providing new insight into the transcriptional programs underlying cell identity in hydrozoans.

ACKNOWLEDGEMENTS

We thank members of the C.E.J., B.H., O.S., and A.D.B. laboratories for valuable research discussions; Ruta Sahasrabudhe, Oanh Nguyen, Diana Burkart-Waco, and Lutz Froenicke from the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) for technical advice and assistance with library preparation and sequencing; Bruce Draper for assistance with imaging; Anh-Dao Nguyen for assistance with generating the genome portal; Vijay Ramani for insight into metazoan chromatin organization; and Birte Mertens for assistance in establishing long-read sequencing in *H. oligactis*. This work was supported by a National Institutes of Health (NIH) grant R35 GM133689 (to C.E.J.), the National Science Foundation: NSF EDGE 1829158 (to C.E.J.), an Austrian Science Fund (FWF) and Deutsche Forschungsgemeinschaft (DFG) DACH

grant I4353 (to O.S.), an Austrian Science Fund (FWF) grant P30347 (to P.L.), and a Fonds National de la Recherche Luxembourg grant 13569708 (to P.B.). This work was also supported in part by the Intramural Research Program of the National Human Genome Research Institute, NIH (ZIA HG000140 to A.D.B.).

AUTHOR CONTRIBUTIONS

J.F.C., S. Siebert, B.H., and C.E.J. designed research; J.F.C., S. Siebert, H.M.L., P.B., A.S.P., M.A., P.L., and O.S. performed research; J.F.C., H.M.L., M.T.F., R.T.M., S. Singh, S.Z., T.G.W., C.E.S., and A.D.B. contributed new analytic tools/resources; J.F.C., S. Siebert, P.B., A.S.P., and O.S. analyzed data; and J.F.C. and C.E.J. wrote the paper with input from O.S., B.H., S. Siebert, and A.D.B.

MATERIALS AND METHODS

Data Availability

We have generated a new genome portal, available at research.nhgri.nih.gov/HydraAEP/, that allows users to interact with and download the data generated in this study. A BLAST server is available to search for genes of interest in the *H. oligactis* and strain AEP *H. vulgaris* gene models. The portal includes an interactive genome browser for visualizing gene models, repetitive regions, ATAC-seq and CUT&Tag peaks, ATAC- seq and CUT&Tag read density, and sequence conservation across the AEP assembly. The website also features an interactive ShinyCell portal for viewing the AEP- aligned *Hydra* single-cell atlas. Step-by-step descriptions of all computational analyses conducted as part of this study, including all relevant code,

formatted both as markdown and HTML documents are available in Supplemental Code S1 and at github.com/cejuliano/brown_hydra_genomes.

The raw sequencing data and assembled genomic sequences data generated in this study have been submitted to the NCBI BioProject database (*https://www.ncbi.nlm.nih.gov/bioproject/)* under accession number PRJNA816482. Note that the chromosome numbering for the version of the strain AEP *H. vulgaris* assembly available via GenBank (Accession JALDPZ000000000) was changed to be consistent with the numbering used for the strain 105 *H. vulgaris* assembly (Accession JAGKSS000000000) (Simakov et al., 2022). We have also made all raw sequencing reads, scripts, and processed data files associated with this study available for download through the genome portal at research.nhgri.nih.gov/HydraAEP/download/index.cgi?dl=fa.

Due to data loss, we no longer have access to the basecall quality scores for the PacBio sequencing data. Because SRA requires that all submitted sequencing data include quality scores, we were unable to upload the PacBio data to NCBI. However, the PacBio data is available at research.nhgri.nih.gov/HydraAEP/download/index.cgi?dl=fa, and the basecall quality scores are not necessary for fully reproducing the results presented in this study.

Hydra strains and animal care

All *Hydra* strains were cultured using standard methods (Lenhoff and Brown, 1970). The AEP strain of *H. vulgaris* was generated from a cross between the PA1 strain isolated by Dr. Carolyn Teragawa from a pond on the Haverford College campus near Philadelphia, Pennsylvania and the CA7 strain isolated by Drs. Lynne Littlefield and Carolyn Teragawa at

Boulder Creek, near Susanville, California (Martin et al., 1997). The DNA used for generating the strain AEP *H. vulgaris* assembly was isolated from a clonally propagated line (the "Kiel" AEP line; courtesy of Thomas Bosch) that was generated from a self-cross of the original AEP line. The DNA used for generating the *H. oligactis* assembly was isolated from the Innsbruck female12 strain, a clonally propagated line originating from a single polyp collected from Lake Piburger See in Tyrol, Austria.

In addition to the Kiel AEP strain, the following lines were used for generating RNA-seq libraries: a transgenic line with an actin::EGFP transgene integrated into the ectodermal lineage and an actin::DsRed2 transgene integrated into the endodermal lineage ("watermelon" line) (Glauber et al., 2013), a transgenic line with an actin::DsRed2 transgene integrated into the ectodermal lineage and an actin::EGFP transgene integrated into the endodermal lineage ("inverse watermelon" line) (Glauber et al., 2013), a transgenic line with an EF1α::EGFP transgene integrated into the endodermal lineage ("enGreen1" line; courtesy of Rob Steele and Catherine Dana), and a transgenic line with a transgene containing EGFP and DsRed2 in an operon configuration with expression driven by the actin promoter integrated into the ectodermal lineage ("operon" line) (Dana et al., 2012).

Hydra vulgaris **strain AEP genome sequencing**

To generate high molecular weight (HMW) genomic DNA (gDNA) libraries for sequencing and assembling the strain AEP *H. vulgaris* genome, we used thirty whole adult polyps from a clonally propagated population belonging to the Kiel AEP line as input. The tissue was flash frozen in liquid nitrogen and HMW gDNA was purified using a Qiagen Gentra Puregene kit

following standard manufactures instructions for mouse tail tissue (Qiagen Cat # 158445; Hilden, Germany). We then performed a Phenol/Chloroform purification using 5PRIME Phase Lock Gels (Quantabio Cat # 2302830; Beverly, Massachusetts) and precipitated the DNA by adding 0.4X 5M ammonium acetate and 3X ice cold ethanol. The DNA pellet was washed twice with 70% ethanol and resuspended in elution buffer (10mM Tris, pH 8.0). We used a Pippin Pulse gel electrophoresis system (Sage Sciences, Beverly, MA) to verify the DNA integrity and a NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts) to verity the DNA purity.

To generate the Oxford Nanopore library, HMW gDNA was gently sheared to 70kb-100kb using a Megaruptor 2 (Diagenode Cat # B06010002; Denville, New Jersey) and the library was prepared using the Oxford Nanopore Ligation Sequencing Kit (Oxford Nanopore Technologies Cat # LSK-109; Oxford, United Kingdom) following standard manufacturer's instructions except for extended incubation times for DNA damage repair, end repair, ligation, and bead elution. 850ng of the final library was loaded on PromethION R9.4.1 flow cells and the data were collected for sixty-four hours. Basecalling was performed live during the run with guppy v1.8.1. A HMW gDNA PacBio library was generated using a SMRTbell *Express* Template Prep Kit 2.0 (PacBio Cat # 100-938-900; Menlo Park, California) following standard manufacturer's instructions. The library was then sequenced on a PacBio Sequel II sequencer using a 1Mv3 SMRT Cell (PacBio Cat # 101-531-000).

To generate the 10X chromium library, HMW gDNA was loaded onto a Chromium Genome Chip (10X Genomics Cat # 120257; Pleasanton, California) and the library was prepared using

Chromium Genome Library & Gel Bead Kit v.2 (10X Genomics Cat # 120258) and Chromium Controller (10X Genomics Cat # 120270) according to manufacturer's instructions with one modification. Briefly, gDNA was combined with Master Mix, Genome Gel Beads, and partitioning oil to create Gel Bead-in-Emulsions (GEMs) on a Chromium Genome Chip. The GEMs were isothermally amplified and barcoded DNA fragments were recovered for Illumina library construction. The post-GEM DNA was quantified using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent Cat # 5067-4626; Santa Clara, California). Prior to Illumina library construction, the GEM amplification product was sheared on an E220 Focused-Ultrasonicator (Covaris Cat # 500239; Woburn, MA) to approximately 375 bp (50 seconds at peak power = 175, duty factor = 10, and cycle/burst = 200). Then, the sheared GEMs were converted to a sequencing library following the 10X standard operating procedure. The library was quantified by qPCR with a Kapa Library Quant kit (Roche Cat # 07960140001; Basel, Switzerland) and sequenced on a HiSeqX10 (Illumina, San Diego, CA) using 2 x 150 bp reads.

For generating the Hi-C library, we used 10 whole flash frozen adult polyps as input. The library was generated using the Arima Hi-C Kit (Arima Genomics Cat # A510008; San Diego, California) following the standard manufacturer's protocol for small animal tissue with the following modification: the frozen tissue was ground using a mortar and pestle for 1 minute in fixation buffer and was subsequently left for 19 minutes at room temperature. The proximally-ligated DNA was fragmented using Covaris E220 (Covaris Cat # 500239) and the biotinylated fragments were enriched. NGS library was prepared using KAPA Hyper prep kit (Roche Cat # 07962363001) and the library was sequenced on an Illumina NovaSeq 6000 using 2 x 150 bp reads.

Whole-animal RNA-seq

To aid in annotating and benchmarking our AEP genome assembly, we generated and sequenced several whole-animal RNA-seq libraries using multiple strain AEP-derived lines. In total, there were 13 libraries: one from the watermelon line, one from the inverse watermelon line, one from the enGreen1 line, one from the operon line, three from male Kiel AEP polyps, three from female Kiel AEP polyps, and three from Kiel AEP polyps that were not producing gametes. For the watermelon, inverse watermelon, enGreen1, and operon RNA-seq libraries, total RNA was purified using a standard Trizol extraction protocol. RNA-seq libraries were then prepared using a TruSeq stranded mRNA kit (Illumina Cat # RS-122-2201) according to the manufacturer's recommended protocol with the following modifications: the RNA was sheared for only 1.5 minutes and the resulting fragments were size selected using a LabChip XT DNA 750 (PerkinElmer Cat # 760541; Waltham, Massachusetts) to be ~500 bp prior to the final PCR enrichment step. The libraries were then sequenced on an Illumina HiSeq2000 using 2 x 100 bp reads.

For the Kiel AEP libraries, total RNA was purified using a standard Trizol extraction protocol. Contaminating DNA was then removed by performing a DNAse digest using the QIAGEN DNAse set (QIAGEN Cat # 79254). A final purification was then performed using the Zymogen RNA Clean and Concentrator Kit (Zymo Research Cat # R1017; Irvine, California) according to the standard manufacturer's protocol. RNA-seq libraries were then generated using the Kapa mRNA-seq Hyper kit (Kapa Biosystems Cat # KK8581; Kapa Biosystems, Cape Town, South Africa). The libraries were then sequenced on a HiSeq4000 using 1 x 50 bp reads. We also performed additional sequencing for one biological replicate from both the male and female Kiel

AEP libraries, which were sequenced on an Illumina HiSeq4000 using 2 x 150 bp reads. To perform the alignment benchmarking analysis presented in Figure B.10, the single end Kiel AEP RNA-seq reads were first processed with Trimmomatic (Bolger et al., 2014) to remove stretches of low-quality base-calls and contaminating adapter sequence. The data was then aligned to both the strain AEP and strain 105 *H. vulgaris* genome assemblies using the RSEM (Li and Dewey, 2011) implementation of STAR (Dobin et al., 2013). The code for this alignment benchmarking analysis is included in the supplemental file 03_aepGenomeAnnotation.md.

Hydra vulgaris strain AEP genome assembly

A step-by-step description of the strain AEP *H. vulgaris* genome assembly methodology, including all relevant code, is provided in the markdown document *01_aepGenomeAssembly* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

The initial draft assembly was generated from the Oxford Nanopore data using Canu (Koren et al., 2017). We then mapped the 10X linked-read data to the draft genome and polished the assembly using Pilon (Walker et al., 2014). For this and all subsequent steps involving the 10X data, we used the 10X Long Ranger pipeline for genome alignment. Following the polishing step, we cut contigs in predicted mis-assembled regions with Tigmint (Jackman et al., 2018) using the 10X data. We then used the 10X data to identify and collapse duplicated contigs in the assembly using Purge Haplotigs (Roach et al., 2018).

Deduplicated contigs were scaffolded with ARCS (Yeo et al., 2018) using the 10X data, and gaps introduced by the scaffolding were filled with PBJelly (English et al., 2012) using the Oxford Nanopore and PacBio data. To generate pseudo- chromosome scaffolds, we aligned the Hi-C data using Juicer (Durand et al., 2016) and scaffolded the assembly using the 3d-dna pipeline (Dudchenko et al., 2017). We subsequently discarded any sequence fragments that were not incorporated into the pseudochromosome scaffolds, as they made up a negligible fraction of the total assembly size (~2.3% of the total assembly sequence) However, these unincorporated fragments are available via the Genbank entry for the AEP genome assembly (accession JALDPZ000000000) for researchers interested in these more difficult to assemble regions. This was followed by an additional gap-filling step with PBJelly using the Oxford Nanopore and PacBio data. To finalize the assembly sequence, we performed another round of Pilon error correction using the 10X, PacBio, and Oxford Nanopore data. Minimap2 (Li, 2018) was used for aligning the long-read data to the genome for the Pilon correction.

The resulting assembly is 901 Mb in length and contains 15 pseudo-chromosome scaffolds, consistent with the haploid chromosome number in *Hydra* (Rahat et al., 1985; Zacharias et al., 2004). Like the strain 105 *H. vulgaris* genome assembly, the AEP assembly is roughly 20- 25% smaller than empirical genome size estimates (~1.06-1.22 Gb for the AEP strain) (Chapman et al., 2010; Zacharias et al., 2004), which is likely due to intrinsic difficulties in resolving long and repetitive stretches of heterochromatin. Nonetheless, the contiguity and completeness of the AEP assembly is comparable to the best currently available hydrozoan genomes (Figure B.1 B and Table B.1). Compared to the recently updated chromosome-level assembly of the strain 105 *H. vulgaris* genome (Simakov et al., 2022), the AEP assembly contains ~10% more sequence (900.9

Mb, compared to 819.4 Mb in the 105 v3 assembly) and a similar number of intact single-copy orthologs predicted from genomic sequence using BUSCO (866, compared to 862 in the 105 v3 assembly; Table S1).

Genome repeat annotation

A step-by-step description of the repeat annotation methodology used for the *H. oligactis* and *H. vulgaris* genomes, including all relevant code, is provided in the markdown document *02_repeatMasking* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To compensate for the lack of well-annotated repeat families available for *Hydra*, we used RepeatModeler2 (Flynn et al. 2020) to predict repeat families ab initio for the *H. oligactis* and strain AEP *H. vulgaris* genome assemblies. We used RepeatMasker (repeatmasker.org) to identify repetitive regions in the strain AEP and strain 105 *H. vulgaris* genome assemblies as well as the *H. oligactis* assembly. For masking repeats in the strain AEP and strain 105 *H. vulgaris* genome assemblies, we used both the strain AEP *H. vulgaris* RepeatModeler2 repeats as well as the Dfam eumetazoan repeat database as repeat libraries when running RepeatMasker. For masking repeats in the *H. oligactis* genome assembly, we used both the *H. oligactis* RepeatModeler2 repeats as well as the Dfam eumetazoan repeat database as repeat libraries when running RepeatMasker. We then used utility scripts included with RepeatMasker to calculate sequence divergence for predicted repeat instances and to generate the repeat landscape plots presented in Figure B.7.

Consistent with previous characterizations of brown *Hydra* genomes, we find that the AEP genome is highly A/T rich (~72%) and repetitive (Chapman et al., 2010; Wong et al., 2019). We estimate that ~71% of the AEP genome is repetitive, with ~6% being simple/low-complexity repeats and ~65% originating from transposable elements (TEs) (Figure B.2 A-C). These estimates are slightly higher than the strain 105 genome (~57% TEs and Figure B.7 D-F) (Chapman et al. 2010). As with the 105 strain, class II TEs—particularly the hAT, CMC, and Mariner families—make up most TE sequences in the AEP genome, although a sizable minority are derived from L2 and CR1 LINE retrotransposons (Figure B.7 A).

Hydra vulgaris strain AEP genome gene annotation

A step-by-step description of the strain AEP *H. vulgaris* genome gene annotation methodology, including all relevant code, is provided in the markdown document *03_aepGenomeAnnotation* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

We generated an initial set of gene models for the strain AEP *H. vulgaris* genome using the BRAKER2 gene prediction pipeline (Brůna et al., 2021). As input into the pipeline, we included the AEP genome sequence with all repetitive regions soft-masked, a custom database of metazoan proteomes, and a whole-animal RNA-seq dataset (described in the "Whole-animal RNA-seq" section above) that was aligned to the soft-masked genome using STAR (Dobin et al., 2013). To supplement the BRAKER2 predictions, we designed a custom annotation pipeline that used exonerate (Slater and Birney, 2005) to generate gene models using transcript sequences from a previously published transcriptome (Siebert et al., 2019) and a manually curated database

of *Hydra* transcript sequences from GenBank. We collapsed duplicated/overlapping gene models in the combined BRAKER2 and exonerate gene predictions by selecting the gene model that had the highest alignment score following a BLAST search against the same custom protein database that was used to generate the BRAKER2 predictions. We then filtered out all gene models that had interrupted reading frames, were shorter than 50 amino acids, or were predicted by InterProScan (Blum et al., 2021; Jones et al., 2014) to contain one or more transposase domains. To improve UTR and splice isoform annotations in our gene predictions, we used the Trinity genome-guided assembly pipeline (Grabherr et al., 2011) to generate a transcriptome from the genome-aligned whole-animal RNA-seq data that was originally used as input for the BRAKER2 pipeline. We aligned this transcriptome to the AEP assembly and used this alignment to update the merged exonerate and BRAKER2 gene models with PASA (Haas et al., 2003), resulting in the final set of gene predictions presented in this study.

Our AEP annotation pipeline identified 28,917 protein coding genes that encode 37,784 predicted transcripts. Although the total gene number is ~14% lower than that observed in the 105 assembly annotations, the AEP annotation contains ~12% more complete single-copy orthologs as predicted using BUSCO (Figure B.1 B and Table B.1), demonstrating an improvement in both accuracy and sensitivity. Furthermore, the AEP assembly gene predictions are the first *H. vulgaris* gene models to include UTRs, with ~48% (13,901) of gene models containing 5' UTRs and ~46% (13,183) containing 3' UTRs. Overall, the AEP gene predictions are comparable to our previously published AEP transcriptome in both the number of predicted transcripts and the number of complete single-copy orthologs (Table S1) (Siebert et al., 2019),

suggesting that our gene annotations have largely captured the transcriptomic repertoire of *H. vulgaris*.

To generate functional annotations for the AEP gene models, we performed a BLAST search against the UniProt protein database (The UniProt Consortium, 2021), predicted protein domains using InterProScan, and identified orthologs in 43 other metazoans using OrthoFinder (Emms and Kelly, 2019). The combined results from these annotation analyses are included in Supplemental Data *S*1. All phylogenies presented in this study were generated as part of the Orthofinder analysis. In the case of the species phylogeny presented in Figure B.21, the branch lengths are derived from the Orthofinder analysis, but the tree's topology was rearranged to be consistent with accepted phylogenies. To identify putative TFs in the AEP gene models, we filtered the InterProScan predictions using a custom set of keywords and GO terms related to transcriptional regulation and DNA-binding activity (see 03_aepGenomeAnnotation.md for details; gene IDs of putative TFs listed in Supplemental Data S1).

Hydra oligactis **genome sequencing**

For generating a draft genome for *H. oligactis*, we prepared two HMW gDNA libraries using the Innsbruck female12 strain of *H. oligactis*. For the first library, HMW gDNA was extracted from 10 whole adult polyps using the Circulomics NanoBind BigTissue kit (Circulomics Cat # NB-900-701-01; Baltimore, Maryland) according to the manufacturer's "Dounce" protocol (Circulomics document # EXT-DHH-001) with the following modifications: we used intact animals instead of finely minced tissue, we homogenized the tissue in 500 µl Buffer CT instead of 750 µl, animals were homogenized using a pestle in a 1.5 ml microcentrifuge tube for 2

minutes instead of using a dounce homogenizer, and the homogenate was pelleted at 1500 G instead of 3000 G. We removed short DNA using the Short Read Eliminator (Circulomics Cat # SS-100-101-01) and Short Read Eliminator XS (Circulomics Cat # SS-100-121-01) kits according to the manufacturer's standard protocol and eluted the samples overnight. We prepared the sequencing library using the Oxford Nanopore Ligation Sequencing Kit (Oxford Nanopore Technologies Cat # LSK-109) according to the standard manufacturer's protocol with the modification that the first two 5-minute incubations were extended to be 30 minutes each. The final library was eluted in 26 µl elution buffer and the library was loaded twice onto an Oxford Nanopore MinION sequencer, with DNAse from the Flow Cell Wash Kit (Oxford Nanopore Technologies Cat # EXP-WSH003) being used to remove gDNA carryover between runs.

The second HMW gDNA library was generated as described above with a few modifications. First, 100 instead of 10 whole animals were used as input. We also made additional modifications to the NanoBind protocol. We prolonged the proteinase K digestion from 30 minutes to 150 minutes, adding another 10 µl proteinase K and another 75 µl Buffer CLE3 90 minutes into the digestion. We also used 30 µl of RNAse A instead of 20 µl. Instead of using a Nanobind disk for DNA extraction as described in the standard protocol, we used the following approach: the lysate was centrifuged at 10,000 G for 5 minutes at room temperature, the resulting pellet was washed with 400 µl Buffer CW1 and centrifuged at 10,000 G for 5 minutes, the pellet was then washed with 500 µl Buffer CW2 and centrifuged at 10,000 G for 5 minutes, the supernatant was removed and the pellet air-dried for 1 minute, and DNA was eluted in 70 µl Elution Buffer. Short gDNA fragment elimination and library preparation was performed as

described for the first library. The library was eluted in 60 µl and was loaded onto the MinION sequencer a total of five times. The total coverage of all sequencing libraries was ~17X (2.4 million reads with an N50 of 22.7 kb).

Hydra oligactis assembly and annotation

A step-by-step description of the *H. oligactis* genome assembly and gene annotation methodology, including all relevant code, is provided in the markdown document *04_oligactisDraftGenome* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

We generated an initial draft assembly for *H. oligactis* with Flye (Kolmogorov et al., 2019) using reads from the two combined Oxford Nanopore libraries described above. The errors in the assembly were then polished with Medaka (github.com/nanoporetech/medaka) using the Nanopore data. To generate a preliminary set of gene models for the draft assembly, we first used previously published whole-animal RNA-seq data from *H. oligactis* (Rathje et al., 2020; Sun et al., 2020) to generate a de novo transcriptome using Trinity (Grabherr et al., 2011). We then aligned this transcriptome to a repeat-masked version of the *H. oligactis* draft genome using minimap2 (Li, 2018). Finally, we ran the BRAKER2 gene prediction pipeline (Brůna et al., 2021), providing as input the repeat-masked *H. oligactis* genome sequence and the genome-mapped Trinity transcriptome.

The *oligactis* assembly is 1274 Mb in length, or ~88% of the empirically estimated genome size (Zacharias et al., 2004). The assembly is ~51-fold more contiguous than the previously available draft genome for *H. oligactis* (N50 of 274.9 kb, compared to previous N50 of 5.4 kb) and has

~27-fold fewer total contigs (16,314 contigs, compared to 447,335 contigs in the previous assembly; Figure B.1 B) (Vogg et al., 2019b). The new *H. oligactis* draft genome is also more complete, with nearly double the number of intact single-copy orthologs (841, compared to 444 in the previous assembly) (Table S1). The A/T and repeat composition (~72% and ~74% respectively) were similar to *H. vulgaris*, although the *H. oligactis* assembly had a slightly higher abundance of repetitive elements (Figure B.1 A and Figure B.7 G-I). We identified 60,590 genes, which is likely an over-estimation of the genome's genic content given that hydrozoan genomes typically contain between 20,000 and 30,000 genes (Chapman et al., 2010; Hamada et al., 2020; Leclère et al., 2019). Nonetheless, the high BUSCO completeness of these gene models (86.2%) suggests that they accurately capture most of the genic content of the *H. oligactis* genome. Thus, we present the first annotated draft genome of *H. oligactis* that is of comparable quality to other published hydrozoan genomes and suitable for systematic comparative analyses.

ATAC-seq

Whole animal ATAC-seq was performed in triplicate on adult bud-free strain AEP *H. vulgaris* polyps using a previously described protocol (Corces et al., 2017; Siebert et al., 2019). All steps of the ATAC-seq protocol were performed using chilled solutions on ice unless otherwise indicated. For each replicate, 5 whole bud-free adult polyps that had been starved for two days were transferred to a sterile 1.5 ml microcentrifuge tube and briefly washed with 1 ml of *Hydra* dissociation medium (DM) (3.6 mM KCl, 6 mM CaCl2, 1.2 mM MgSO4, 6 mM sodium citrate, 6 mM sodium pyruvate, 6 mM glucose, 12.5 mM TES buffer, adjusted to pH 6.9) (Gierer et al., 1972). The polyps were then homogenized in 1 ml DM using ~50 strokes of a tight-fitting glass dounce. The homogenate was transferred into a sterile 1.5 ml microcentrifuge and spun down at

500 G for 5 minutes in a centrifuge chilled to 4℃. The cell pellet was resuspended in 50µl resuspension buffer (RSB) (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, pH 7.4) containing 0.1% Tween-20, 0.1% NP-40, and 0.01% digitonin. Lysis proceeded for 3 minutes and was subsequently halted by adding 1 ml RSB containing 0.1% Tween-20. Nuclear density in the lysate was quantified by loading 19 µl of the resuspension and 1 ul of 20mM Hoechst 33342 (ThermoFisher Scientific Cat # 62249; Waltham, Massachusetts) onto a Fuchs-Rosenthal hemocytometer. An aliquot of the resuspended lysate containing ~50,000 nuclei was then transferred to a fresh 1.5 ml microcentrifuge tube and was subsequently spun down for 10 minutes at 500 G in a centrifuge chilled to 4℃. The crude nuclear pellet was then resuspended in 50 ul tagmentation buffer (1X TD buffer [Illumina Cat # 20034197], 33% phosphate-buffered saline, 0.01% digitonin, 0.1% Tween-20, 5 ml TDE1 [Illumina Cat # 20034197]) and shaken at 1000 rpm for 30 min at 37℃. Tagmentation was halted by adding 250 µl of PB buffer from a QIAGEN MinElute PCR Purification Kit (QIAGEN Cat # 28004; Hilden, Germany).

Tagmented DNA was purified using a QIAGEN MinElute PCR Purification Kit using the standard manufacturer's instructions. The libraries were eluted in 21 µl water and amplified for an initial five PCR cycles using 2X NEBNext master mix (NEB Cat # M0541S; Ipswitch, MA) following the cycling parameters specified in the original ATAC-seq protocol (Buenrostro et al., 2015, 2013). The number of additional PCR cycles following this initial amplification was then determined by performing qPCR on an aliquot of the pre-amplified libraries as described in the original ATAC-seq protocol. Biological replicate 1 received 1 additional cycle of PCR (for a total of 6), replicate 2 received 3 additional cycles (for a total of 8), and replicate 3 received 4 additional cycles (for a total of 9). Two rounds of post-PCR clean-up were performed using

Agencourt AMPure XP beads (Beckman Coulter Cat # A63881; Pasadena, California) following the standard manufacturer's protocol. During this step we selected for DNA fragments between 100 and 700 bp in size. Library concentration was quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific Cat # Q32851) and fragment size distributions were determined using the Bioanalyzer High-Sensitivity DNA kit (Agilent Cat # 5067-4626). The libraries were then pooled at roughly equimolar proportions and sequenced on an Illumina NextSeq 500 using 2 x 75 bp reads.

CUT&Tag

CUT&Tag targeting H3K4me1, H3K4me3, and H3K27me3 were each performed in triplicate using a modified version of the originally published CUT&Tag protocol (Kaya-Okur et al. 2019) that was adapted for use in *Hydra*. Each CUT&Tag replicate consisted of 40 whole, bud-free strain AEP *H. vulgaris* polyps that had been fed once weekly and then starved for two days prior to the experiment. Unless otherwise specified, all steps were performed at room temperature without agitation. The polyps were collected in a 1.5 ml microcentrifuge tube, washed once with 1 ml DM, and then homogenized in 1 ml DM using 40 strokes of a tight-fitting glass dounce. The homogenate was passed through a 70 μm filter and centrifuged for 5 minutes at 1000 G. The resulting pellet was resuspended in 1 ml of lysis buffer (20mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1X cOmplete protease inhibitor [Roche Cat # 11836153001], 2 mM EDTA, 0.1% tween-20, 0.1% NP-40, and 0.01% digoxygenin) and incubated for 5 minutes. The lysate was centrifuged for 5 minutes at 1300 G to produce a crude nuclear pellet, which was then resuspended in 1 ml of wash buffer (20mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1X cOmplete protease inhibitor) and divided evenly into 4 1.5 ml microcentrifuge tubes. The

volume of each tube was then brought to 1 ml using wash buffer. 10 µl of 5mg/ml Concanavalin A coated magnetic beads (Bangs Laboratories Cat # BP531; Fishers, Indiana) that had first been washed twice in bead activation buffer (20 mM HEPES, pH 7.5, 10 mM KCl, 1 mM CaCl2, and 1 mM MnCl2) was added to each tube. The bead-nuclei suspensions were then incubated for 10 minutes on a rotator and the supernatant was subsequently removed using a magnet stand. Bead-bound nuclei were resuspended in 50 µl solutions of either 1:1000 negative control rabbit IgG (EpiCypher Cat # 13-0042; Durham, North Carolina), 1:100 rabbit α-H3K4me1 (Abcam Cat # ab8895; Cambridge, United Kingdom), 1:100 rabbit α-H3K4me3 (Active Motif Cat # 39060; Carlsbad, California), or 1:50 rabbit α-H3K27me3 (Cell Signaling Technology Cat # 9733T; Danvers, Massachusetts) diluted in antibody buffer (1% bovine serum albumin and 2 mM EDTA in wash buffer). The nuclei were incubated in the primary antibody solutions for 2 hours. This was followed by a 1-hour incubation in 50 µl of anti-rabbit secondary antibodies (EpiCypher Cat # 13-0047) diluted 1:100 in antibody buffer. The nuclei were then quickly washed three times in 1ml wash buffer, resuspended in 50 µl of 1x pAG-Tn5 (EpiCypher Cat # 15-1017) diluted in high-salt buffer (20mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1X cOmplete protease inhibitor), and incubated for 1 hour. Next, excess pAG-Tn5 was removed using three quick 1 ml washes with high-salt buffer and the nuclei were resuspended in 150 µl of tagmentation buffer (high-salt buffer with 10 mM MgCl2 added). Tagmentation was then allowed to proceed for 1 hour at 37℃. Tagmentation was stopped by adding 5 µl 0.5 mM EDTA, 1.5 µl 10% SDS, and 2.5 µl proteinase K (ThermoFisher Scientific Cat # EO0492) to each sample and incubating at 55℃ for 1 hour.

Tagmented DNA was purified using a Zymogen Oligo Clean & Concentrator Kit (Zymo Research Cat # D4060; Irvine, California) following the standard manufacturer's protocol. The libraries were eluted in 21 µl water and amplified using 2X NEBNext master mix following the cycling parameters described in the original CUT&Tag protocol (Kaya-Okur et al. 2019) for a total of 13 cycles. We then used Agencourt AMPure XP beads to perform two rounds of post-PCR clean-up and to select for DNA fragment sizes between 100 and 700 base pairs. We quantified the concentration of our libraries using the Qubit dsDNA HS Assay Kit and we determined their fragment size distributions using the Bioanalyzer High-Sensitivity DNA kit. When measuring the concentrations of our purified libraries, we found that our negative control samples were too dilute to effectively validate their size and concentration for pooling. We therefore performed another five rounds of PCR amplification on the three negative control libraries followed by two additional rounds of AMPure bead cleanup. Finally, libraries were pooled at roughly equimolar concentrations and sequenced on an Illumina NextSeq 500 using 2 x 75 bp reads.

Cis-regulatory element annotation

A step-by-step description of the *Hydra cis*-regulatory element annotation methodology, including all relevant code, is provided in the markdown document *08_creIdentification* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To analyze the ATAC-seq data collected from whole strain AEP *H. vulgaris* polyps, we first filtered the raw reads using Trimmomatic (Bolger et al., 2014) to remove stretches of low-

quality base-calls and contaminating adapter sequence. The filtered reads were then aligned to the AEP assembly using Bowtie2 (Langmead and Salzberg, 2012). To remove mitochondrial reads, we also aligned the ATAC-seq data to the *Hydra* mitochondrial genome (Voigt et al., 2008) and subsequently discarded any reads that aligned to the mitochondrial and nuclear genome references using Picard Tools (broadinstitute.github.io/picard/). We next identified and removed PCR duplicates from the aligned data using Samtools (Li et al., 2009) and Picard Tools. We then called peaks for each ATAC-seq biological replicate using MACS2 (Zhang et al., 2008) (Zhang et al. 2008). To generate a consensus peakset of biologically reproducible ATAC-seq peaks, we first calculated irreproducible discovery rate (IDR) (Li et al. 2011) peak scores for each pairwise combination of biological replicates (three in total). We defined a reproducible peak as one that received an IDR score ≤ 0.1 for at least two pairwise comparisons between biological replicates. Transcription factor binding footprints were predicted using TOBIAS (Bentsen et al., 2020).

We identified 50,151 ATAC-seq peaks, 12,807 H3K4me1 peaks, 1,969 H3K4me3 peaks, and 3,744 H3K27me3 peaks (Supplemental Data S3). The number of ATAC-seq peaks we identified in the AEP assembly is similar to previously published *Hydra* ATAC-seq datasets generated using strain 105 animals (Cazet et al., 2021; Siebert et al., 2019). However, the number of peaks from our CUT&Tag libraries likely underrepresent the true number of genomic regions enriched for each respective histone modification.

Thus, although we have demonstrated for the first time that CUT&Tag can successfully be applied to a cnidarian model, the protocol will require further optimization to improve sensitivity

in the future. The establishment of CUT&Tag in *Hydra* offers substantial benefits over alternative chromatin mapping techniques, namely ChIP-seq, as CUT&Tag requires approximately two orders of magnitude fewer animals as input compared to equivalent *Hydra* ChIP-seq experiments (Reddy et al., 2020).

To analyze the CUT&Tag data collected from whole strain AEP *H. vulgaris* polyps, we first used Trimmomatic to remove stretches of low-quality base-calls and contaminating adapter sequence. We then aligned the data to the AEP assembly using Bowtie2. PCR duplicates were then identified and removed using Samtools. We then called peaks for the H3K4me1 H3K4me3 and H3K27me3 data with SEACR (Meers et al., 2019) using the IgG data as the background signal. To identify biologically reproducible peaks, we again performed IDR and selected peaks with an IDR score ≤ 0.1 for at least two of the three pairwise comparisons between biological replicates.

We used UROPA (Kondili et al., 2017) to annotate all ATAC-seq and CUT&Tag peaks based on the nearest TSS. We used deepTools (Ramírez et al., 2016) to generate the correlation heatmap globally comparing the aligned CUT&Tag and ATAC-seq data, to generate the data tracks used to depict read density along the AEP assembly, and to characterize the distribution of ATAC-seq and CUT&Tag data in and around genes. Individual plots visualizing the CUT&Tag, ATAC-seq, and sequence conservation data were generated using Gviz (Figure B.1 B; Hahne and Ivanek, 2016) and pyGenomeTracks (Figure B.9 and B.12; (Lopez-Delisle et al., 2020).

Systematically characterizing cnidarian 3D chromatin organization

A step-by-step description of the single-cell RNA-seq atlas mapping and annotation methodology, including all relevant code, is provided in the markdown document *09_3dChromatin* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To characterize chromatin organization in the strain AEP *H. vulgaris* genome, the raw Hi- C reads were re-mapped to the finalized assembly using the Juicer pipeline (Durand et al., 2016). Subsequently, contact matrices were normalized and domain boundaries predicted with the HiCExplorer pipeline (Ramírez et al., 2018) using a bin size of 16 kb. To characterize gene expression patterns around predicted contact domain boundaries, we first identified sets of three genes that spanned predicted contact domain boundaries using bedtools. We then used R to calculate the Pearson correlation score for both gene pairs that either abutted domain boundaries (intra-domain pairs) or spanned domain boundaries (inter-domain pairs) using the NMF gene score values calculated from the *Hydra* single cell atlas (described below in "*Hydra* single-cell atlas mapping and annotation"). A student's T-test, as implemented in R, was used to test for a significant difference in correlation values between inter- and intra-domain gene pairs. Hi-C contacts and domain boundaries were visualized using Juicebox (Robinson et al., 2018).

To systematically characterize the 3D organization of cnidarian genomes, raw Hi-C reads were downloaded from NCBI for the following species: Nematostella vectensis (Zimmermann et al., 2022; PRJNA667495), *Acropora millepora* (Hoencamp et al., 2021; PRJNA512907), *Rhopilema esculentum* (Nong et al., 2020; PRJNA505074), *Haliclystus octoradiatus* (PRJEB45135), and

Diadumene lineata (PRJEB46842). The raw reads were then mapped using the Juicer pipeline. Knight and Ruiz normalized read count matrices for all possible scaffold-to- scaffold pairs were then exported with a bin size of 100 kb using Juicer Tools (Durand et al., 2016). We then quantified inter-centromeric interactions by quantifying the average total number of inter-chromosomal contacts at every position along every pseudo-chromosome scaffold in each assembly. These contact values were then converted to a z-score and the highest z-score found along each scaffold was selected as the 'inter-centromeric contact score'. We then used Tukey's Honest Significant Difference method as implemented in R to perform a post-hoc significance test on an ANOVA calculated on all inter-centromeric contact scores for all species considered in the analysis. We used a significance cutoff of $p \leq 0.05$. To quantify inter-telomeric interactions, we performed an aggregate chromosome analysis (ACA; Hoencamp et al., 2021) using a bin size of 500 kb. Typically, this analysis requires centromere coordinates, but because we only used ACA for quantifying telomere interactions—a calculation that does not depend on accurate centromere coordinates—we simply set these coordinates to be approximately at the midpoint of each pseudo-chromosome scaffold.

Hydra **single-cell atlas mapping and annotation**

A step-by-step description of the single-cell RNA-seq atlas mapping and annotation methodology, including all relevant code, is provided in the markdown document *05_hydraAtlasReMap* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

We aligned the raw *Hydra* single-cell atlas sequencing data (previously deposited under BioProject PRJNA497966) to the AEP genome transcript models using the Drop-seq Tools alignment pipeline (github.com/broadinstitute/Drop-seq). Following mapping, we next determined which cell barcodes to include in downstream analyses. Because most beads in a Drop-seq experiment are not exposed to a lysed cell, only a small minority of sequenced cell barcodes are associated with a genuine single-cell transcriptome. Instead, most barcodes have low read counts attributable to contamination from ambient RNA. To differentiate between cell barcodes containing true single-cell transcriptomes and barcodes containing only transcriptomic noise, we generated plots that depicted the cumulative read fraction of cell barcodes ordered by read depth from highest to lowest. The curves generated by these plots have an elbow—an inflection point where the cumulative read fraction rapidly plateaus. This inflection point demarcates the transition from true biological signal to noise. For our downstream analyses, we used only read count data from the cell barcodes that preceded the elbow in the cumulative read plot.

Subsequent clustering and visualizations of the scRNA-seq data were done using Seurat (Hao et al., 2021b). Prior to clustering, we performed additional filtering to remove cell barcodes with fewer than 300 or greater than 7,500 unique molecular identifiers (UMIs) as well as barcodes with fewer than 500 or greater than 75,000 total reads. We also removed any genes that were found in fewer than 3 cells. After filtering, we normalized the data using sctransform (Hafemeister and Satija, 2019) and corrected for batch effects using reciprocal PCR as implemented in Seurat. We then clustered the single-cell transcriptomes using the Louvain algorithm (Waltman and Van Eck, 2013) and visualized the results using a UMAP dimensional

reduction (McInnes et al., 2018). We annotated the clustered dataset using a panel of previously validated marker genes (Figures B.17 and B.18 A) (Siebert et al., 2019).

As with prior analyses of the *Hydra* scRNA-seq atlas (Siebert et al., 2019), we found that many individual cell transcriptomes simultaneously contained multiple transcripts known to have mutually exclusive expression patterns. These chimeric transcriptomes are referred to as "doublets" and can result from either technical or biological causes (Evan Z. Macosko et al., 2015; Siebert et al., 2019). For example, battery cells, a prominent source of doublets in *Hydra* scRNA-seq data, are tentacle ectodermal cells in which both neurons and nematocytes are stably embedded (Bode and Flick, 1976; Hufnagel et al., 1985; Yu et al., 1985). Because these three cell types are tightly physically associated in battery cell complexes, they are resistant to dissociation and are frequently sequenced as a single cell (Siebert et al., 2019).

To systematically identify likely doublets, we identified markers associated with ectodermal, endodermal, neuronal, nematocyte, gland, and germ cells using a Wilcoxon Rank Sum test as implemented in Seurat. We then calculated a holistic score representing how highly each cell in the atlas expressed each set of cell type markers using the Seurat AddModuleScore function (Figure B.18 B). Because most doublets in *Hydra* include at least one epithelial cell (Siebert et al. 2019), we defined a doublet as a cell with a score greater than 0.2 for both an epithelial module and any other cell type module (Figure B.18 C). For the sake of clarity and simplicity, we chose to exclude all doublet transcriptomes from the finalized version of the AEP genome-mapped atlas; however, we provide an alternative version of the atlas with doublets included (available at research.nhgri.nih.gov/HydraAEP/download/index.cgi?dl=fa), as certain cell types

(e.g., battery cells) may require the inclusion of doublets to be properly represented. We repeated the batch correction, clustering, and UMAP dimensional reduction after removing all predicted doublets and found two remaining clusters, one that contained endodermal/interstitial doublets and another that appeared to contain cells expressing stress markers, that we removed prior to finalizing the set of cells included in the doublet-free version of the atlas. We again repeated the clustering and UMAP dimensional reduction steps to generate the final atlas presented in the main text, which we annotated using the same panel of previously validated marker genes described above.

To identify groups of co-expressed genes in the single-cell atlas, we performed non- negative matrix factorization (NMF) as implemented in the cNMF python package (Kotliar et al., 2019) on the full (doublets included) single-cell expression matrix. NMF is a dimensional reduction technique that, when applied to gene expression data, groups co-expressed genes into modules referred to as metagenes. The number of metagenes identified by NMF, a value referred to as k, needs to be specified prior to performing the factorization. The optimal value for k cannot be determined objectively and instead needs to be estimated empirically by evaluating a range of k values. Therefore, we performed an initial parameter sweep using k values ranging from 15 to 90 by steps of 5. The results from NMF depend on how the analysis is initialized, so we performed 200 independent runs for each k value that could then be combined to generate a consensus factorization result. We then selected a k value that maximized reproducibility across independent runs while simultaneously minimizing the differences between the factorized data and the original expression data. Based on these criteria, we selected a k value of 55. Our initial sweep of k values used steps of 5, so to more precisely identify the optimal k value we

performed another parameter sweep for k values ranging from 50 to 60 by steps of 1. After evaluating the reproducibility and fidelity of the results from the fine resolution sweep, we selected a final k value of 56. We then generated the final consensus factorization results after first discarding individual runs that contained irreproducible results (see 05_hydraAtlasReMap.md for details).

In situ hybridization

To generate labeled RNA probes for performing in situ hybridization, we cloned and sequenced PCR products for the *Hydra* genes *G017021* (*parascleraxis*) and *G008733* that had been amplified from oligo-dT-primed cDNA generated from whole adult male and female *H. vulgaris* polyps (Kiel AEP line). Amplicons were generated using the following PCR primers: *G017021*-forward: AGTTTAAAATGCTCCAATCTATAAGG; *G017021*-reverse: TAATACGACTCA CATAGGGTGATCTTAAAAATGTAACGCAAAATG; *G008733*-forward: GCTTTAGGCGGCTCAA CAAA; *G008733*-reverse: ATTTAGGTGACACTATAGAACCTTTGTTTACGCCAGCA. The reverse primer sequences for *G017021* and *G008733* included T7 and SP6 promoter sequences respectively, allowing us to use purified PCR products as templates for in vitro transcription reactions using the Roche DIG RNA Labeling Kit (Roche Cat # 11175025910). The resulting DIG- labeled RNA products were then purified using the Zymogen RNA Clean & Concentrator-25 kit (Zymo Research Cat # R1017) and stored at -80℃ until use.

To perform in situ hybridization on whole *Hydra* polyps, we used a slightly modified version of a previously published protocol (Bode et al. 2009). For each in situ, 15 whole adult strain AEP

H. vulgaris polyps that had been starved for two days were transferred to a 1.5 ml microcentrifuge tube, relaxed at room temperature (RT) for 1 minute in 1 ml *Hydra* medium (HM) containing 2% urethane, and then fixed in 1 ml HM containing 4% paraformaldehyde (PFA) at 4℃ overnight. All subsequent steps were performed at RT in 1 ml of solution with gentle rocking agitation unless otherwise indicated. Following overnight fixation, the fixative was removed with three quick washes in PBT (0.1% tween-20 in phosphate buffered saline, pH 7.4). The tissue was then bleached by transferring the samples gradually to 100% MeOH using 5-minute washes first in 33% MeOH in PBT then in 66% MeOH. The samples were then incubated in 100% MeOH for 1 hour. To maximize bleaching, the samples were incubated overnight in fresh 100% MeOH at - 20℃. The tissue was rehydrated using 1 wash with 66% MeOH, 1 wash with 33% MeOH in PBT, and three washes in PBT for 5 minutes each. The tissue was then permeabilized in 10 µg/ml proteinase K in PBT for 5 minutes. Proteinase activity was halted with a quick wash in 4 mg/ml glycine in PBT followed by a 10-minute wash in fresh glycine solution. The glycine solution and any residual proteinase K was then removed with three 5-minute washes in PBT. The samples were then washed twice in 0.1 M triethanolamine in PBT, once in 0.1 M triethanolamine in PBT containing 3 µl/ml acetic anhydride, once in 0.1 M triethanolamine in PBT containing 6 µl/ml acetic anhydride, then three times in PBT, all for 5 minutes each. Next, the tissue was refixed for 1 hour using 4% PFA in PBT. The fixative was removed with three 5-minute PBT washes followed by two 5-minute washes in 2X SSC (300 mM NaCl and 30 mM sodium citrate). In preparation for probe hybridization, the samples were incubated in 50% 2X SSC/50% hybridization solution (HS; 50% formamide, 5x SSC [750 mM NaCl and 75 mM sodium citrate], 1x Denhardt's solution, 100 µg/mL heparin, 0.1% Tween-20, and 0.1% Chaps) for 10 minutes, starting first at RT then gradually transitioning to hybridization

temperature (56℃). All subsequent pre-hybridization and hybridization steps were carried out at 56℃. The tissue was incubated in HS for 10 minutes and then in HS containing 200 µg/ml yeast RNA for 2 hours. To prepare the DIG-labeled probes for hybridization, we added ~750 ng of probe to modified HS (50% formamide and 5x SSC) and denatured secondary RNA structures by incubating the solution at 85℃ for 5 minutes. The probe solution was then added to the sample tubes after first being diluted in fresh HS containing 200 µg/ml yeast RNA to a final probe concentration of ~3 ng/ul. The samples were then left to hybridize for ~60 hours with no agitation.

Excess probe was removed using a sequence of single, 5-minute washes in HS, 75% HS/25% 2X SSC, 50% HS/50% 2X SSC, and then 25% HS/75% 2X SSC at 56℃. The samples were then washed twice with 2X SSC containing 0.1% CHAPS for 30 minutes each, with the first wash occurring at 56˚C and the second at 37˚C. Unbound probe was digested by treating the tissue with 20 µg/ml RNase A in 2X SSC containing 0.1% CHAPS for 30 minutes at 37℃ without agitation. RNase A was then removed using two 10-minute washes at 37℃ and two 30-minutes washes at 55℃ in 2X SSC containing 0.1% CHAPS. The samples were then transitioned back to RT and washed three times with MABT (100 mM maleic acid, 150 mM NaCl, 0.1% Tween 20, pH 7.5) for five minutes each. Non-specific protein interactions in the tissue were then blocked with a two-hour incubation in blocking solution (MABT containing 1% BSA and 20% sheep serum) at 4℃. The samples were then resuspended in a 1:2000 dilution of Anti-Digoxigenin-AP (Roche Cat # 11093274910) in blocking solution and incubated overnight at 4℃ without agitation.

Following antibody binding, the samples were transitioned back to RT and excess antibodies were removed with eight 20-minute washes in MABT. The tissue was then washed once in NTMT (100 mM NaCl, 100 mM Tris-HCl, 50 mM MgCl2, 0.1% Tween-20, pH 9.5) for 5 minutes. During this NTMT wash, the samples were transitioned to six-well plates. The NTMT was then replaced with 20 µl/ml NBT/BCIP solution (Roche Cat # 11681451001) in NTMT. The staining reaction proceeded for an empirically determined time (~1-2 hours) and was subsequently stopped using three quick PBT washes. To reduce non-specific signal in the tissue, the tissue was transitioned into 100% EtOH using 5-minute washes first in 33% EtOH in PBT then in 66% EtOH. The tissue was then incubated in 100% EtOH until the staining in the tissue turned from purple to blue (~30 minutes). The tissue was then rehydrated using single 5-minute washes in 66% EtOH then 33% EtOH in PBT. Finally, residual EtOH was removed using three quick PBT washes. The in situs were documented using a Leica DM5000B microscope (camera Leica DFC310FX), a Leica M165C digital stereo microscope (camera MC170HD), or a Zeiss Axiophot microscope (camera Leica DFC 550).

Characterization of gene age in the *Hydra* single-cell atlas

A step-by-step description for our methodology for characterizing the cell-type-specific transcriptional patterns associated with gene age, including all relevant code, is provided in the markdown document *06_geneAge* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To estimate the age for each *Hydra* gene model, we adopted a phylostratigraphic approach (Domazet-Lošo et al., 2007). We used the orthology predictions generated from our OrthoFinder

analysis (see "AEP genome gene annotation") to identify the most recent clade that contained all orthologs of each *Hydra* gene (i.e. the "clade of origin"). We defined gene age to be the age of each gene's clade of origin. For example, if a gene in *Hydra* had orthologs throughout Cnidaria, but lacked any orthologs outside of Cnidaria, then Cnidaria would be considered that gene's clade of origin. Therefore, the gene likely first emerged after the split of Bilateria and Cnidaria but before the split of Anthozoa and Medusozoa.

We next used these gene age predictions to characterize the relationship between gene age and cell-type specific transcription in our *Hydra* single-cell atlas. To do this, we first generated lists of genes that were present in the transcriptomes of each cell type in our atlas by identifying all genes with an average expression level above 0.05 normalized counts per cell for each cell type. Then, to exclude ubiquitously expressed genes that do not vary across different cell types, we used the Seurat FindVariableFeatures function to identify 7,500 genes with high or intermediate levels of variability across the *Hydra* atlas and excluded genes from our cell type transcriptomic profiles if they were not found in this variable gene list. To calculate the relative enrichment of each age across *Hydra* cell types, we calculated the odds that a gene expressed in a certain cell type will be of a certain age. We found that the transcriptomes of all cell types were heavily skewed towards ancient genes that predate Metazoa, likely reflecting the essential and deeply conserved functions of ancient genes. However, cell-type-specific enrichment patterns did emerge when we normalized the enrichment profiles across cell types by scaling the values in each column to have a maximum value of 1 and a minimum value of 0. We calculated single-cell transcriptomic age index values by applying a previously described formula (Domazet-Lošo and Tautz, 2010) to the normalized *Hydra* atlas single-cell gene expression matrix.

Whole-genome alignment and phylogenetic footprinting

A step-by-step description of the single-cell RNA-seq atlas mapping and annotation methodology, including all relevant code, is provided in the markdown document *07_genomeConservation* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

We generated a cross-species whole-genome alignment of the *C. hemisphaerica*, *H. viridissima*, *H. oligactis*, strain 105 *H. vulgaris*, and strain AEP *H. vulgaris* genome assemblies using Progressive Cactus (Armstrong et al. 2020). To facilitate the alignment, we ensured that repetitive regions in each genome were soft-masked. In the case of the *Clytia* and *H. viridissima* genomes, we made use of publicly available repeat-masked data. Repeats in *H. oligactis*, strain AEP *H. vulgaris*, and strain 105 *H. vulgaris* were masked with RepeatMasker using repeat families identified by the RepeatModeler2 pipeline (Flynn et al., 2020).

To quantify sequence conservation rates in across the AEP assembly using the resulting alignment, we used a custom Python script to count the number of non-AEP genomes with the same nucleotide for every position of the AEP assembly that was included in the whole-genome alignment. For visualizing the sequence conservation results (as in Figure B.2 B), we smoothed the per- base conservation results using a 100 bp moving window. We used deepTools (Ramírez et al., 2016) to characterize the distribution of conservation rates around the AEP assembly gene models.

To identify putative conserved transcription factor binding sites (TFBS) in the AEP assembly, we first used FIMO (Grant et al., 2011) to identify putative binding sites in all four *Hydra* genomes in our alignment using a custom database of non-redundant vertebrate, insect, and nematode binding motif sequences from the JASPAR database (Fornes et al., 2020). To generate a control dataset, we also performed TFBS prediction using a version of our custom motif sequence database where the nucleotide order of each motif had been shuffled. We then used the Hierarchical Alignment API (Hickey et al., 2013) in conjunction with our cross-species genome alignment to convert the coordinates of all non-AEP TFBS coordinates to their equivalent coordinates in the AEP assembly. This allowed us to determine if a given TFBS in the AEP assembly was also present in other *Hydra* genomes. We considered a TFBS in the AEP assembly to be conserved if it was present in the strain 105 *H. vulgaris* assembly and at least one other *Hydra* genome. To further filter our conserved TFBS list to sites that were most likely to be functionally relevant, we eliminated any predicted binding sites that did not fall within an ATAC- seq peak or that overlapped protein coding sequence. To identify motif sequences from our custom database that showed evidence of conservation in *Hydra*, we used a chi-square test, as implemented in R, to identify motifs with significantly (FDR ≤ 0.01) higher conservation rates than shuffled controls.

To identify putatively conserved CREs, we used deepTools (Ramírez et al., 2016b) to calculate the average level of sequence conservation for each ATAC-seq and CUT&Tag peak in the AEP assembly. We calculated these sequence conservation rates using pairwise comparisons between the AEP assembly and each non-AEP assembly in our whole-genome alignment, such that each peak received four separate conservation scores (e.g., one score for the AEP-105 alignment, one

score for the AEP-*oligactis* alignment, etc.). We then used k-means clustering, as implemented in R, to partition peaks into two populations—a high-scoring population and a low-scoring population—for each pairwise species comparison. We defined a peak as conserved if it was classified as high scoring in at least two pairwise comparisons. To characterize the distribution of conserved enhancer-like CREs around genes in the AEP- assembly (presented in Figure B.2 C,D), we used UROPA (Kondili et al., 2017) to calculate the distance from each H3K4me1 and ATAC-seq peak to the nearest TSS. To remove possible core promoter peaks from this analysis, we disregarded all H3K4me1 and ATAC-seq peaks that overlapped a H3K4me3 peak prior to visualizing the TSS distance distribution.

To perform syntenic analyses, we used D-GENIES to generate whole-genome alignments and corresponding dotplots using the strain AEP *H. vulgaris* genome as a target sequence and the strain 105 *H. vulgaris*, *H. viridissima*, and *H. oligactis* genomes as queries (Cabanettes and Klopp, 2018). Within the D-GENIES application, minimap2 (v. 2.24) was used for generating the alignment using the "Many Repeats" repeatedness configuration of D-GENIES. Spurious alignments were removed from the resulting dotplots using the "Hide Noise" function.

Prediction of transcriptional regulators in *Hydra*

A step-by-step description of the *Hydra* transcriptional regulator analysis, including all relevant code, is provided in the markdown document *10_hydraRegulators* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To identify motifs enriched in the putative regulatory regions of genes belonging to cell- type-specific gene co-expression programs in the *Hydra* single-cell atlas, we used gene set enrichment analysis (GSEA) as implemented in the fgsea R package (Korotkevich et al., 2021; Subramanian et al., 2005). GSEA requires two inputs: 1) a set binary of classifications that groups together genes associated with a feature or process of interest (i.e., a gene set), and 2) a set of continuous scores that can be used to rank genes. To test for enrichment, GSEA evaluates if the members of a given gene set show a non-random distribution in their score rankings (i.e., if the gene set is biased towards having higher or lower scores). If a gene set has a non-random distribution, it indicates that the feature or process that was used to group those genes (e.g., the presence of a specific motif in nearby regulatory regions) is associated with the metric used to generate the gene rankings (e.g., a gene co-expression score for a specific cell type). The strength of this association is quantified using a metric called the normalized enrichment score, with higher scores indicating a stronger bias for the gene set to be associated with high gene ranks.

To perform a motif enrichment analysis using GSEA, we used our conserved TFBS predictions (described above in "Whole-genome alignment and phylogenetic footprinting") to generate gene sets that grouped genes according to to the conserved binding motifs that were present in their putative regulatory regions, such that each motif was assigned a list of genes that were predicted to be regulated by the motif's cognate TF. For the continuous scores used to order genes in the GSEA, we used the *Hydra* atlas NMF gene scores (NMF described in "Single- cell atlas mapping and annotation"), which reflect how strongly the expression pattern of a gene matched the expression pattern associated with a given metagene. After performing GSEA for each metagene in the *Hydra* atlas, we discarded any enrichment scores that were not significant (adjusted P-

value > 0.01) to reduce noise in the enrichment results. We then mapped these enrichment scores onto the *Hydra* atlas by generating single-cell enrichment scores for each motif. To do this, we used NMF cell scores, which reflect how well each metagene reflected a cell's overall transcriptomic profile, to calculate a weighted average enrichment score for each cell, with enrichment scores from highly scoring metagenes contributing more strongly than lowly scoring metagenes.

To identify the candidate transcription factors that could plausibly bind the motifs associated with each metagene, we first used metadata available through the JASPAR and UniProt databases to identify the Pfam DNA-binding domains present in each motif's cognate TF. We then generated a list of candidate regulators for each motif by identifying the AEP gene models that possessed the appropriate DNA-binding motifs. To determine the most likely candidate regulators for each motif, we used the single-cell atlas to identify TFs whose expression was correlated with the enrichment pattern of their cognate motif.

A common problem that arises when performing correlation analyses using single cell RNA-seq data is the high frequency of 'dropouts', instances where low and moderately expressed genes are completely missed in a random subset of cell transcriptomes due to low sequencing depth. To mitigate this source of noise, and thus facilitate the comparison of motif enrichment and TF expression patterns, we used the *Hydra* atlas NMF results to generate an imputed version of the single-cell expression data. The results of a single-cell RNA-seq NMF analysis are two matrices, a gene score matrix and a cell score matrix, that approximate the original expression matrix when

multiplied together. This NMF-derived approximation eliminates the cell-to-cell heterogeneity caused by dropouts, thus facilitating single-cell expression correlation analyses.

Using the imputed read count matrix, we performed a correlation analysis to identify motifs whose enrichment pattern was correlated with the expression pattern of a TF that possessed the appropriate DNA-binding domain. TFs with a motif enrichment correlation score ≥ 0.5 were deemed candidate regulators. We also reviewed possible regulator/motif pairs manually, allowing us to catch marginal cases where TFs were expressed in only a subset of cells where the target motif was enriched, causing them to fall slightly below our correlation score threshold (e.g., *zic1* and *zic4*). The final selection of the motif/TF pairs we presented in the figures of this study were selected manually from the list of candidates generated by the systematic analysis described above.

To control for the possible contribution of sequence bias to our enrichment results, we repeated our GSEA and TF expression correlation analysis using shuffled versions of each motif (see 10_hydraRegulators.md for details). We found that while some shuffled motifs were significantly enriched in the *Hydra* atlas, the enrichment patterns of the shuffled motifs were overwhelmingly different from the enrichment patterns of their unshuffled counterparts. Specifically, the enrichment patterns of over 90% (832/907) of shuffled motifs had a correlation score ≤ 0 when compared to the enrichment patterns of the unshuffled motifs. This demonstrates that the enrichment patterns we observed using the unshuffled motifs were not driven primarily by sequence composition biases. We also found that the correlation scores between motifs and their candidate regulators were significantly higher when using unshuffled motifs when

compared to shuffled motifs (student's t-test P-value ≤ 2.2e-16), suggesting the enrichment patterns for the unshuffled motifs better reflected the regulatory activity of *Hydra* TFs.

Re-aligning the *Clytia* single-cell atlas

A step-by-step description of the approach for generating new *Clytia* gene models and the subsequent re-alignment and clustering of the *Clytia* single-cell atlas, including all relevant code, is provided in the markdown document *11_clytiaAtlasReMap* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

The initial published version of the *Clytia* single cell RNA-seq atlas used a newly generated set of gene models for the original version of the *Clytia* genome as a reference for read mapping (Chari et al., 2021; Leclère et al., 2019). However, we used an updated version of the *Clytia* genome (available at metazoa.ensembl.org/Clytia_hemisphaerica_gca902728285) for our cross-species whole genome alignment. To maintain a consistent genome reference across analyses, and to maximize the completeness of the gene models used for mapping the single cell data, we generated a custom set of gene predictions for the updated version of the *Clytia* genome. To do this, we first generated a preliminary set of gene predictions by aligning both the new transcriptome generated in the *Clytia* single-cell atlas publication and the transcript models from the original *Clytia* genome publication to the updated Clytia genome using PASA. We then combined the PASA gene models with the gene models for the updated genome assembly using AGAT (github.com/NBISweden/AGAT). The resulting gene models were more complete than the pre-existing gene models for the updated genome assembly, as indicated by the increased

number of complete single copy orthologs identified using BUSCO (Table S1). We then aligned the raw *Clytia* single-cell data to the newly generated transcript models using the 10X Cell Ranger pipeline. Following mapping, we selected the cell barcodes used for downstream analysis by retaining only those cells that were present in the original published version of the *Clytia* atlas. We then clustered the re-mapped data using the Louvain algorithm as implemented in Seurat and found that our analysis recapitulated the cell type clustering results from the original publication (see 11_clytiaAtlasReMap.md for details), validating our mapping and clustering approach.

To characterize the cell-type-specificity of *Clytia* genes that were lost in the *Hydra* lineage, we first used the results from our OrthoFinder analysis (described above in "AEP genome gene annotation") to identify *Clytia* genes with orthologs in *Hydractinia echinata* (the other non-*Hydra* hydrozoan in our analysis) but with no orthologs in any of the *Hydra* proteomes in our analysis. We then generated a holistic score representing how strongly each cell in the *Clytia* atlas expressed these lost genes using the Seurat AddModuleScore function.

Aligning the *Clytia* and *Hydra* single-cell atlases

A step-by-step description of the *Clytia* and *Hydra* single-cell RNA-seq atlas alignment, including all relevant code, is provided in the markdown document *12_crossSpeciesAtlasAlignment* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To align the *Clytia* and *Hydra* single cell atlases, we first identified all *Hydra* genes with unambiguous one-to-one orthologs in *Clytia* using the results from our OrthoFinder analysis (described above in "AEP genome gene annotation"). We then subset the *Clytia* and *Hydra* single-cell read count matrices to only include these one-to-one orthologs and converted all *Clytia* gene names to their *Hydra* equivalent. After the data was reformatted, we used reciprocal principal component analysis as implemented in Seurat to combine and align the *Hydra* and *Clytia* single-cell RNA-seq data. We then performed Louvain clustering on the aligned data and visualized the results using a UMAP dimensional reduction. We annotated the resulting clusters by propagating the cell type annotations associated with each cell barcode from the uncombined versions of the *Clytia* and *Hydra* atlases.

To quantify the transcriptional similarities between *Clytia* and *Hydra* cell types, we made use of a previously described alignment metric (Tarashansky et al., 2021). To calculate this alignment score, we performed a mutual nearest neighbor analysis (MNN) as implemented in the BiocNeighbors R package. This analysis identified all cross-species cell pairs where each member of the pair was among the other's 30 nearest cross-species neighbors in principal component space. We calculated the alignment score by determining the portion of total MNNs for a cell type of interest that belonged to each cell type in the other species. We retained all cross-species cell type pairs with an alignment score ≥ 0.05. We also calculated a single-cell divergence score, which measures the average distance between a cell and it's thirty nearest cross-species neighbors in principal component space. A smaller divergence score thus indicates that the transcriptomic profile of a given cell is more like the transcriptomic profiles of cells from the other species than cells with higher divergence scores.

To identify genes with conserved expression patterns in *Clytia* and *Hydra*, we first performed a high-resolution Louvain clustering analysis to generate 'pseudo-cells' that grouped together small sets of *Clytia* and *Hydra* cells with similar gene expression profiles. We then calculated average gene expression values for each species in each pseudo-cell. We designated a gene as having a conserved expression pattern if the pseudo-cell expression values in the two species had a correlation score > 0.65.

Predicting conserved transcriptional regulators in *Clytia* and *Hydra*

A step-by-step description of the *Clytia* transcriptional regulator analysis and the comparison of candidate regulator predictions in *Hydra* and *Clytia*, including all relevant code, is provided in the markdown document *13_conservedRegulators* available at github.com/cejuliano/brown_hydra_genomes. This document is also provided in Supplemental Code S1.

To identify cell-type-specific gene co-expression modules in *Clytia*, we performed NMF on the raw *Clytia* atlas single-cell expression matrix, following the same steps as described above for the *Hydra* single-cell atlas (see "Single-cell atlas mapping and annotation"). To identify the optimal number of metagenes, we first performed a broad sweep of k values from 15 to 90 by steps of 5. We observed a local maximum in the stability of the NMF results for k=40, prompting us to perform a second sweep of k values from 35 to 45 by steps of 1. Based on this fine resolution sweep, we chose a k value of 37. We then generated the final consensus factorization results after first discarding individual runs that contained irreproducible results.

Because *cis*-regulatory element annotations were not available for *Clytia*, we were unable to use the same motif enrichment approach as for our analysis in *Hydra*. Instead, to isolate presumptive promoter sequences we extracted all sequences that fell within 1 kb upstream of a TSS. Then, for each *Clytia* metagene, we generated a ranked list of these putative promoters with sequences that were near genes strongly associated with the metagene placed at the top of the list and sequences near genes that were weakly associated placed at the bottom of the list. We then used these ranked promoters as input for an Analysis of Motif Enrichment (AME) (McLeay and Bailey, 2010). To map the AME results onto the *Clytia* single-cell atlas, we calculated single- cell weighted averages of the significant (E-value < 10) fold-enrichment results for each metagene using the NMF metagene cell scores.

To identify conserved regulators in *Hydra* and *Clytia*, we manually reviewed the expression patterns and associated motif enrichment patterns for all TFs that both had a conserved expression pattern in *Clytia* (see Aligning the *Clytia* and *Hydra* single-cell atlases") and were designated as candidate regulators in *Hydra* (see "Prediction of transcriptional regulators in *Hydra*"). We considered a TF to be a conserved regulator when both the expression of the TF and the enrichment of its cognate motif were localized to the same cell populations in *Clytia* and *Hydra* in the cross-species single-cell atlas.

To determine if the degree of overlap in motif enrichment patterns for the *Hydra* and *Clytia* atlases was greater than would be expected by chance, we repeated our analysis using shuffled versions of each motif. We then quantified the degree of overlap in motif enrichment patterns using the same pseudo-cell correlation approach described above (see "Aligning the *Clytia* and

Hydra single-cell atlases"). We observed no highly correlated ($r \geq 0.5$) enrichment patterns when using shuffled motifs, whereas we found 13 highly correlated enrichment patterns when using unshuffled motifs (Supplemental Data S16). This suggests that the enrichment overlap we observed using unshuffled motifs are likely indicative of conserved TF function and are not driven purely by chance.

FIGURES

A

Species	*Hydra vulgaris*	*Hydra vulgaris* (v3)	*Hydra vulgaris* (v2)	*Hydra oligactis*	*Hydra oligactis2*	*Hydra viridissima*
Strain	AEP	105	105	Innsbruck female12	Cold Resistant (CR)	A99
Citation	This study	Simakov et al., 2022	Chapman et al., 2010	This Study	Vogg et al., 2019	Hamada et al., 2020
Total Size	901 Mb	819.4 Mb	854 Mb	1274 Mb	1262 Mb	284 Mb
Scaffolds	15	54	5525	16314	447335	2677
Longest Scaffold	85.9 Mb	78.0 Mb	4.4 Mb	6.4 Mb	380 kb	5.1 Mb
Scaffold N50	58.6 Mb	55.0 Mb	1.0 Mb	274.9 kb	5395 bp	1.1 Mb
Scaffold L50	7	7	246	1197	64225	73
Scaffold N90	43.9 Mb	42.2 Mb	161.8 kb	33.9 kb	1145 bp	111.4 kb
Scaffold L90	15	14	984	5740	264614	391
Repeat Composition	72.2%	N/A	64.1%	73.9%	N/A	37.5%
Gaps	561	2207	121961	7	23151	27305
Ambiquous Bases	56.1 kb	2.2 Mb	6.8 Mb	700 bp	771 kb	47.7 Mb
Genes	28917	N/A	33820	60,590	N/A	21467
BUSCO (genome)	90.7%	90.3%	86.1%	88.2%	46.5%	88.0%
BUSCO (protein)	94.7%	N/A	82.8%	86.2%	N/A	88.2%

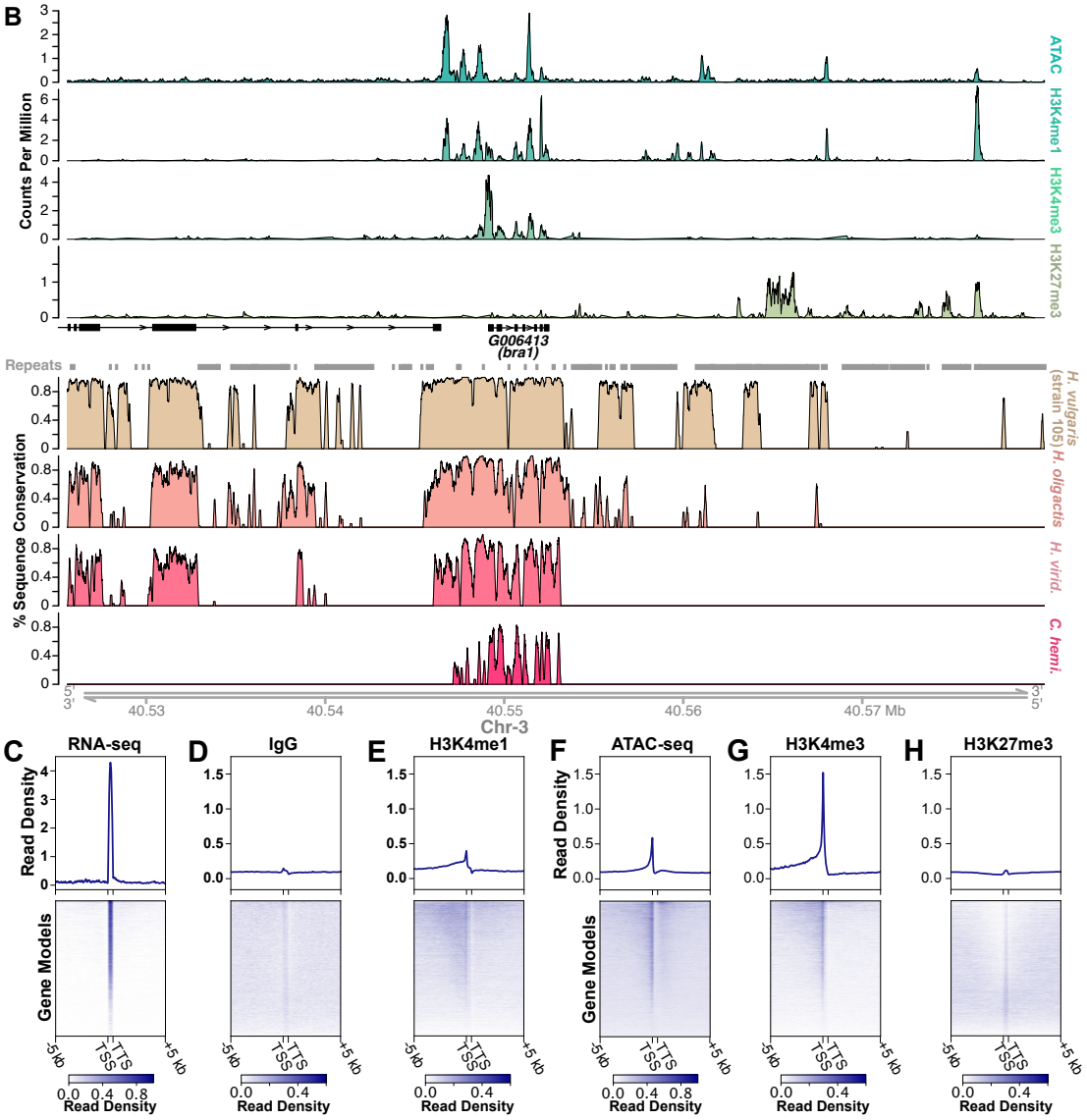


Figure B.1. New genome assemblies provide improved resources for *Hydra* molecular biology research. (A) The *Hydra vulgaris* strain AEP and *Hydra oligactis* genome assemblies presented in this study are marked improvements on the previously available reference genomes

for their respective species. (B) Representative plot of CUT&Tag, ATAC-seq, and genomic conservation tracks centered on the *hybra1* gene (Technau and Bode, 1999). For the sequencing data, each track represents the signal from pooled biological replicates for the specified library type. A plot of the same locus that includes separate tracks for each CUT&Tag and ATAC-seq biological replicate is presented in Supplemental Figure B.9. (C–H) Read distribution for sequencing data centered on AEP assembly gene models. (C) Whole-animal RNA-seq data are strongly enriched in predicted coding sequences. (D) Control IgG CUT&Tag reads show minimal enrichment in or around genes. (E) H3K4me1 is enriched in promoter-proximal regions, but only weakly enriched at transcription start site (TSS). (F) ATAC-seq is enriched at TSS, but also shows some enrichment in more distal regions, likely because ATAC-seq targets both promoters and enhancers. (G) H3K4me3 is strongly enriched at the TSS. (H) H3K27me3 shows minimal or no enrichment near transcribed genes. (TTS) Transcription termination site.

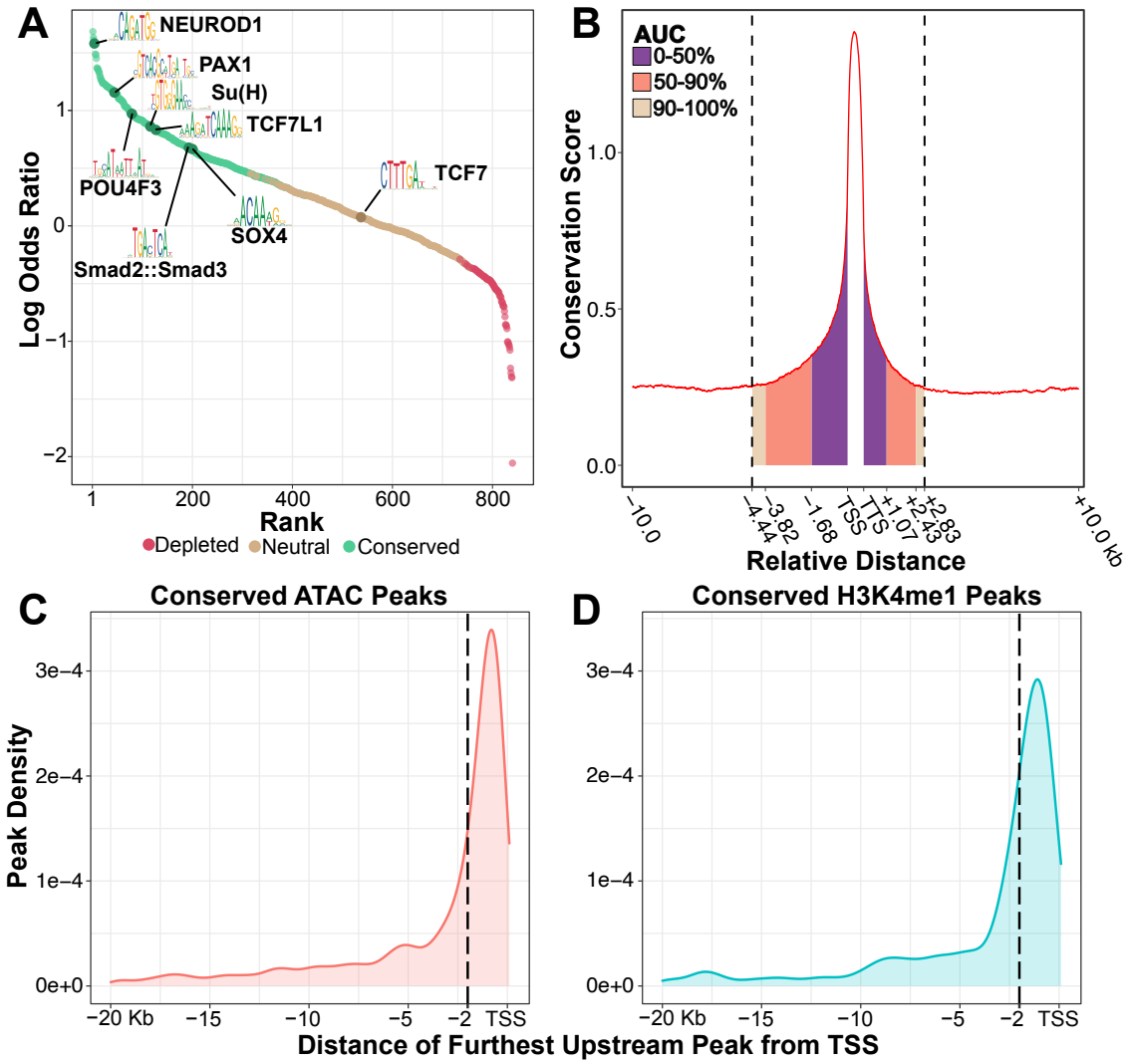


Figure B.2. Phylogenetic footprinting reveals conserved regulatory elements and transcription factor (TF) binding sites across the *Hydra* genome. (A) Quantification of TF binding motif conservation across four *Hydra* genomes. A positive log odds value indicates the nonshuffled motif had a higher conservation rate than its shuffled control. Statistical significance was evaluated using a chi-square test with an FDR cutoff of 0.01. (B) The distribution of sequence conservation levels around genes suggests that a sizable minority of promoter-proximal CREs extends farther than 2 kb from the nearest TSS. The conservation score represents the average number of non-AEP hydrozoan genomes that had the same base as the AEP assembly at a given locus. (AUC) Area under the curve; (TSS) transcription start site; (TTS) transcription termination site. Gene bodies were excluded from the AUC calculation. (C,D) Distribution plots summarizing the distance from the furthest upstream CRE for each gene to the predicted target TSS based on either ATAC-seq (C) or H3K4me1 CUT&Tag (D) Dotted vertical lines demarcate 2 kb from the TSS.

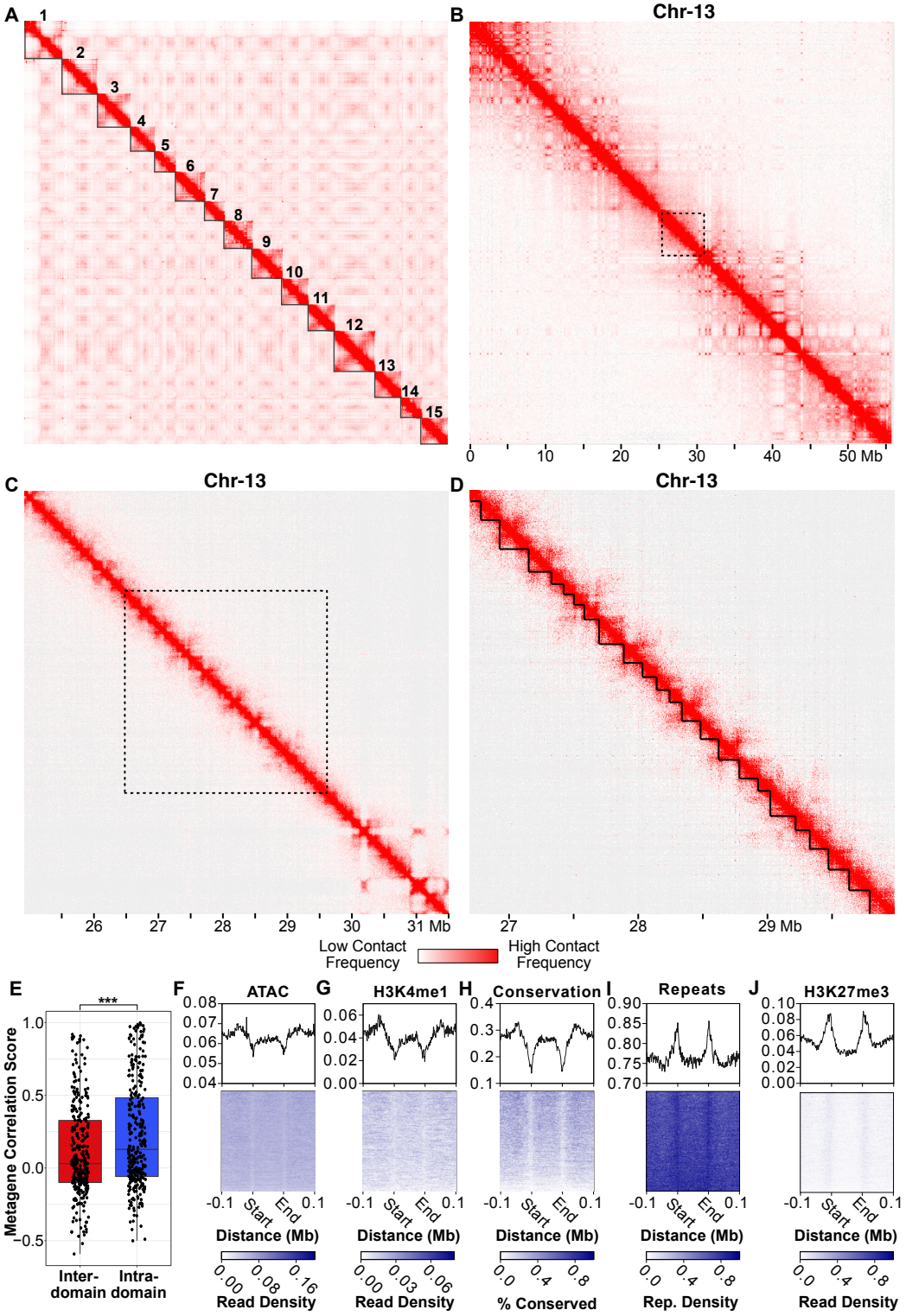


Figure B.3. Hi-C data reveal hierarchical chromatin architecture in the *Hydra* genome. (A) Hi-C contact map for the *H. vulgaris* strain AEP assembly reveals 15 pseudochromosomes with high levels of both inter-chromosomal interactions between presumptive centromeric regions and intra-chromosomal interactions between centromeric and telomeric regions. (B) The chromatin interaction map for Chromosome 13 reveals megabase-scale chromatin

compartments. The black dotted lines indicate the region visualized in the subsequent figure panel. (C) Kilobase-scale interaction domains can be found within a single megabase-scale compartment. (D) Representative depiction of predicted kilobase-scale chromatin interaction domains in *Hydra* (black lines). (E) Boxplot/scatterplot depicting the correlation in expression for adjacent gene pairs show that gene pairs within the same domain (intra-domain pairs) were significantly more similar than pairs that spanned a domain boundary (inter-domain pairs; Welch two-sample *t*-test *P*-value = 6.93×10^{-5}). (F–J) Predicted domain boundaries fall within regions of heterochromatin. Domain boundaries are associated with reduced chromatin accessibility (F), H3K4me1 (G), and sequence conservation (H) and with elevated repeat element density (I) and H3K27me3 (J).

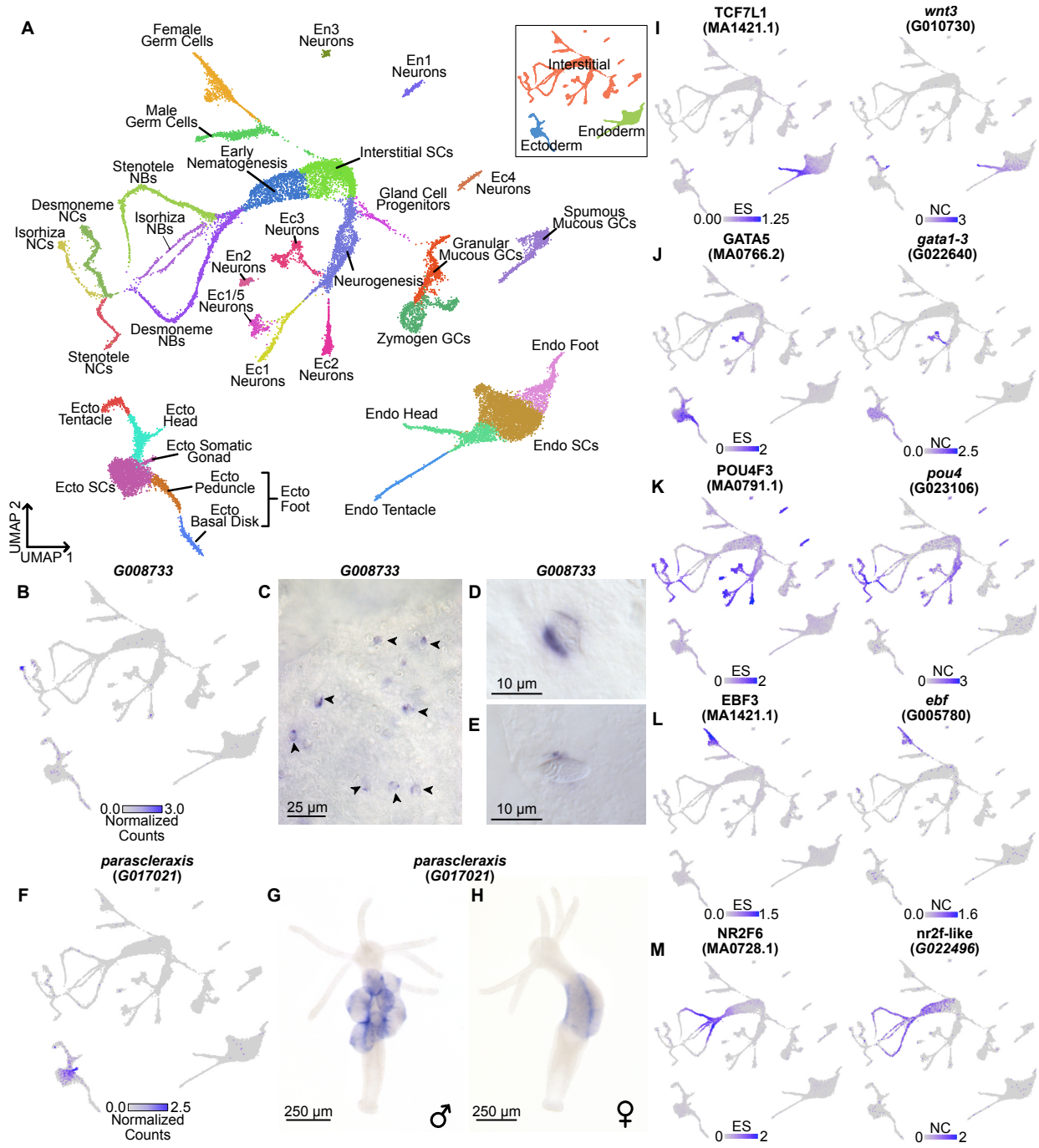


Figure B.4. An updated *Hydra* single-cell RNA-seq atlas reveals novel regulators of gene co-expression in *Hydra*. (A) Uniform manifold approximation and projection (UMAP) dimensional reduction of the *Hydra* single-cell RNA-seq atlas mapped to the AEP reference genome captures virtually all known cell states in adult polyps. *Inset* shows UMAP colored by the three stem cell lineages in adult *Hydra*. (NCs) Nematocytes; (NBs) nematoblasts; (SCs) stem cells; (Ecto) ectodermal epithelial cells; (Endo) endodermal epithelial cells; (GCs) gland cells; (Ec) neuron subtypes found in the ectoderm; (En) neuron subtypes found in the endoderm. (B) The gene *G008733* is a specific marker for isorhiza nematocytes. (C–E) In situ hybridization targeting *G008733* labels isorhiza nematocytes (black arrowheads) in upper body column tissue.

(F) *scleraxis* is a specific marker for ectodermal somatic gonad cells. (G) In situ hybridization targeting *scleraxis* in male polyps labels ectodermal testes cells. (H) In situ hybridization reveals *scleraxis* is expressed in egg patches in female polyps. (I–M) Motif enrichment and gene expression patterns reveal candidate regulators of cell state. (I) TCF motif enrichment and *wnt3* expression data corroborate the role of TCF/Wnt signaling in epithelial head tissue. (J) GATA motif enrichment and expression data corroborate the role of *gata1-3* in aboral epithelial tissue and suggest an additional function in Ec3 neurons. (K) Pou4 motif enrichment and expression data suggest *pou4* regulates transcription in differentiating and mature neurons and nematocytes. (L) Ebf motif enrichment and expression data suggest *ebf* regulates transcription during oogenesis. (M) NR2F motif enrichment and expression data suggest *nr2f-like* regulates transcription during nematogenesis. Corresponding JASPAR motif IDs are provided in the parenthetical text under the motif names (Fornes et al., 2020). (ES) Enrichment score; (NC) normalized counts.

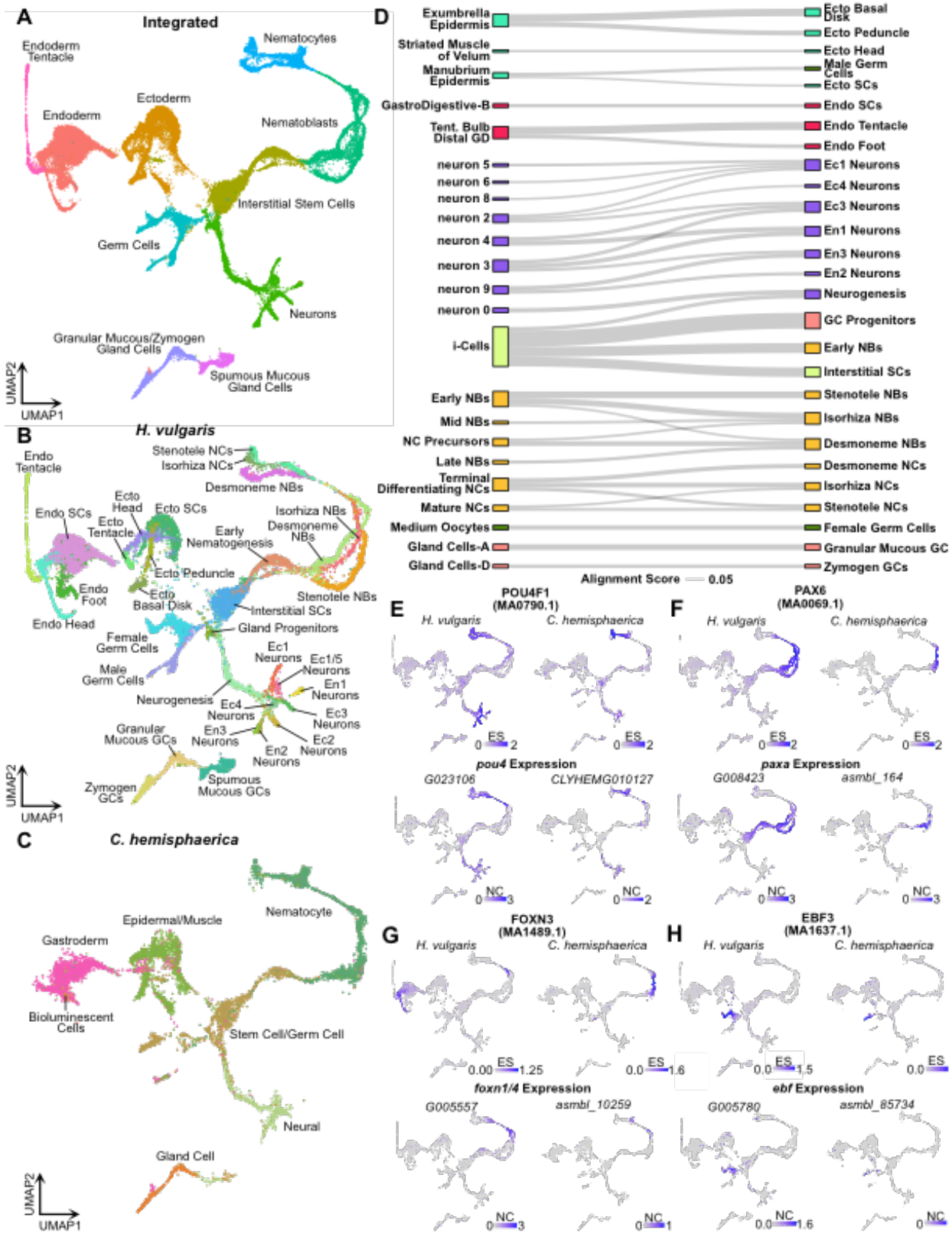


Figure B.5. Aligned *Hydra* and *Clytia* single-cell atlases reveal conserved cell type–specific transcriptional regulation. (A–C) UMAP dimensional reduction of aligned *Hydra* and *Clytia* medusa single-cell atlases clusters together equivalent cell types from the two species. (D)

Sankey plot showing transcriptional similarities between *Hydra* (*right* column) and *Clytia* (*left* column) cell types highlights extensive similarities among interstitial cell types. The alignment score quantifies the proportion of mutual nearest neighbors for one cell type that are made up of members of another cell type. An alignment score threshold of 0.05 was used to exclude poorly aligned cell types. (NCs) Nematocytes; (NBs) nematoblasts; (SCs) stem cells; (Ecto) ectodermal epithelial cells; (Endo) endodermal epithelial cells; (GCs) gland cells; (Ec) neuron subtypes found in the ectoderm; (En) neuron subtypes found in the endoderm; (Tent.) tentacles; (GD) gastroderm. (E–H) Conserved motif enrichment and gene expression patterns reflect gene regulatory network conservation in hydrozoans. (E) *pou4* is a conserved regulator of late stage and mature neurons and nematocytes. (F) *paxA* is a conserved regulator of nematoblasts. (G) *foxn1/4* is a conserved regulator of nematocyte maturation. (H) *ebf* is a conserved regulator of oogenesis. (ES) Enrichment score; (NC) normalized counts.

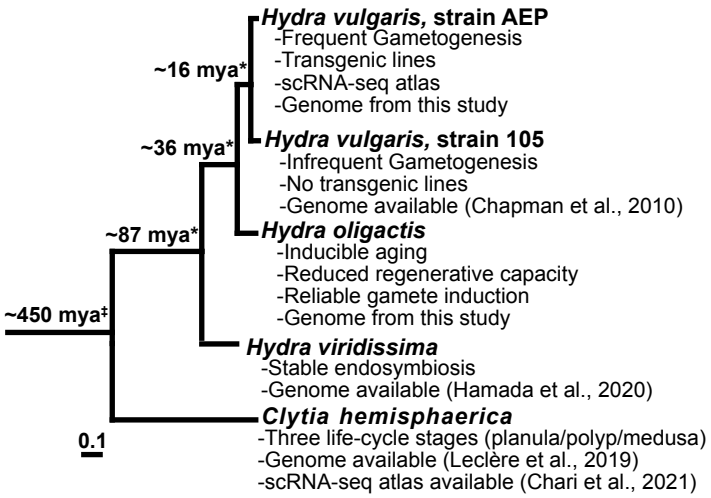


Figure B.6. Phylogeny of hydrozoan research organisms highlighting currently available genomic and transcriptomic resources, divergence time estimates, and evolutionary gains and losses. * indicates divergence time estimates taken from Wong et al. (2019). ‡ indicates divergence time estimate taken from Dohrmann and Worheide (2017).

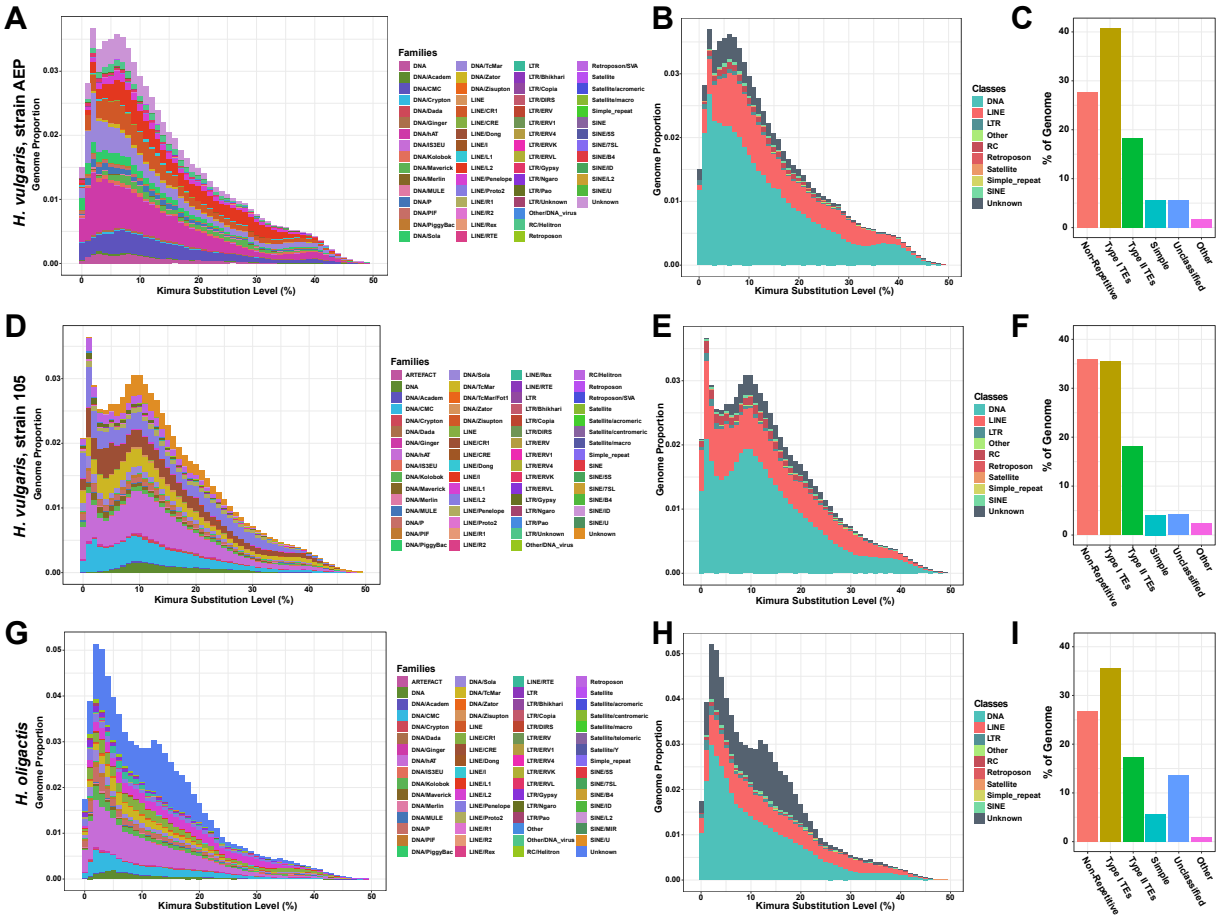


Figure B.7. Repeat composition of *Hydra* genomes. Summary plots of repeat composition in the (A-C) strain AEP *H. vulgaris*, (D-F) strain 105 *H. vulgaris*, and (G-I) *H. oligactis* genomes. Repeat landscapes are presented at the level of repeat subfamilies (A, D, and G) and broader repeat classes (B, E, and H). (C, F, and I) Total proportions for repetitive and non-repetitive elements across each genome.

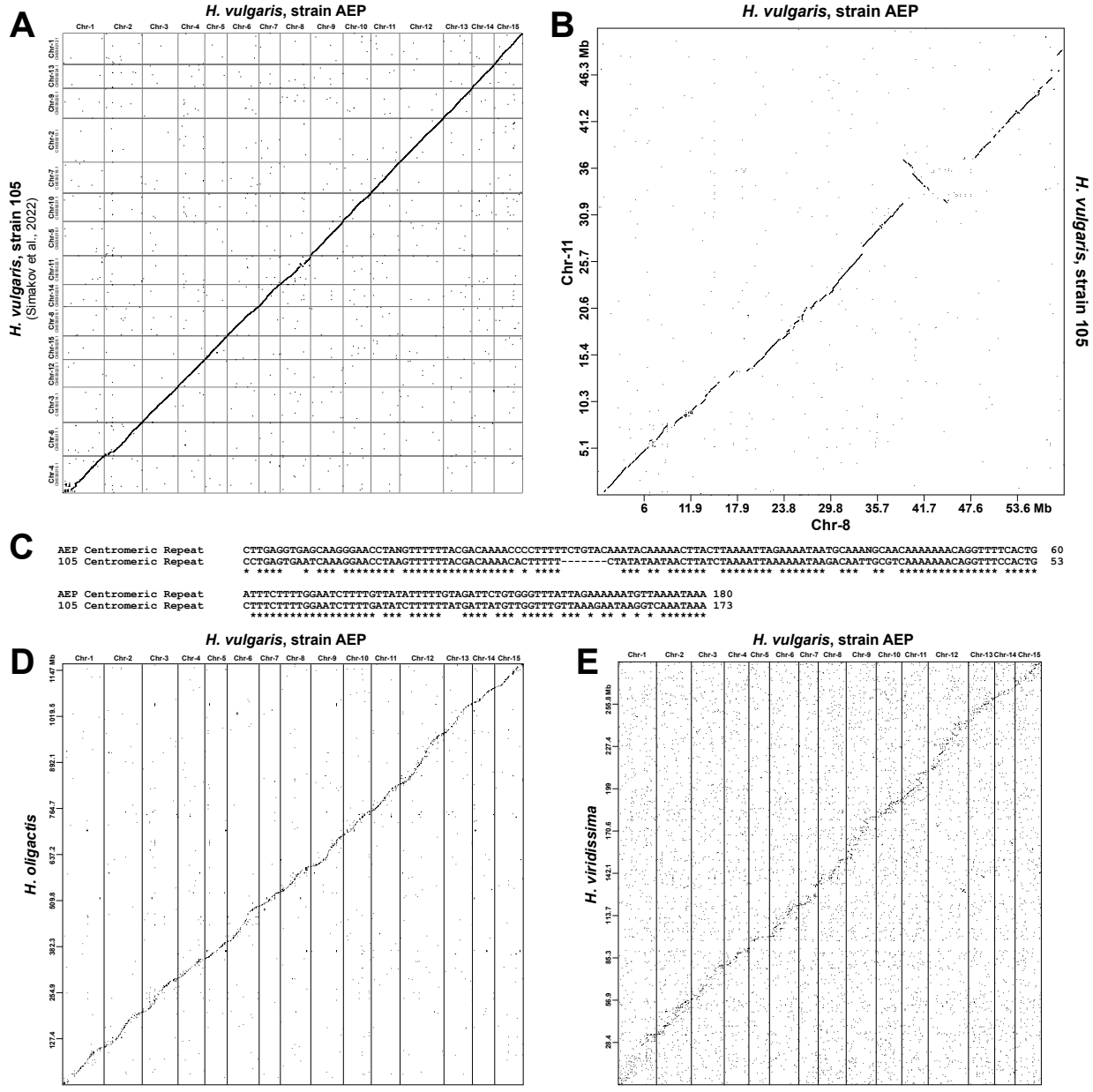


Figure B.8. Comparative analysis of *Hydra* genome sequences. (A) Dotplot reveals highly conserved synteny from the strain 105 *H. vulgaris* genome assembly to the strain AEP assembly. (B) There has been a ~5 Mb inversion on chr-8 since the split of strain 105 and strain AEP of *H. vulgaris*. (C) Alignment of the centromeric repeats from the strain AEP and strain 105 *H. vulgaris* genomes (Melters et al. 2013). (D,E) Preliminary chromosome assignments for the (D) *H. oligactis* and (E) *H. viridissima* genome assemblies based on synteny dotplots.

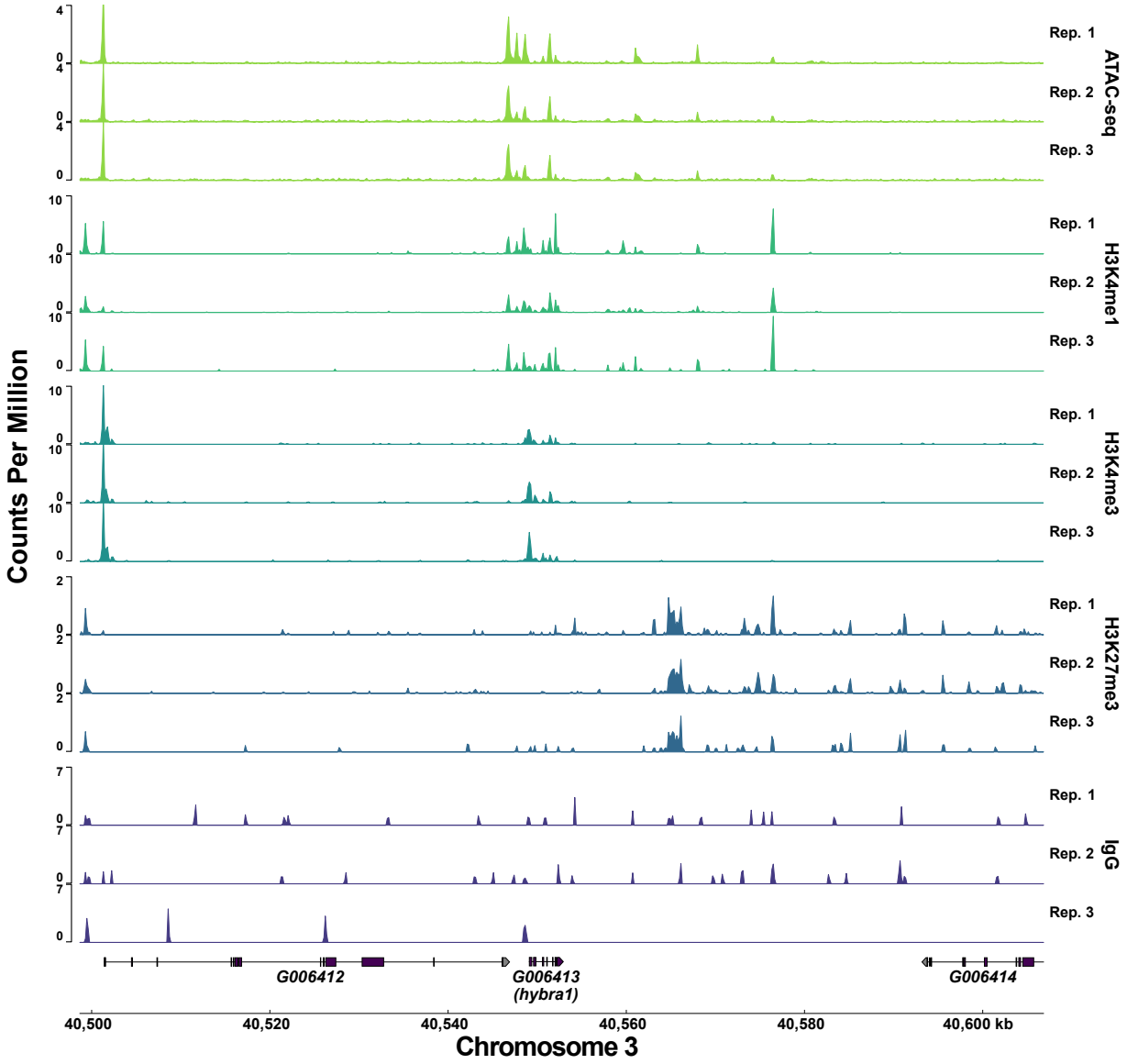


Figure B.9. Representative plot of all CUT&Tag and ATAC-seq biological replicates centered on the *hybra1* gene.

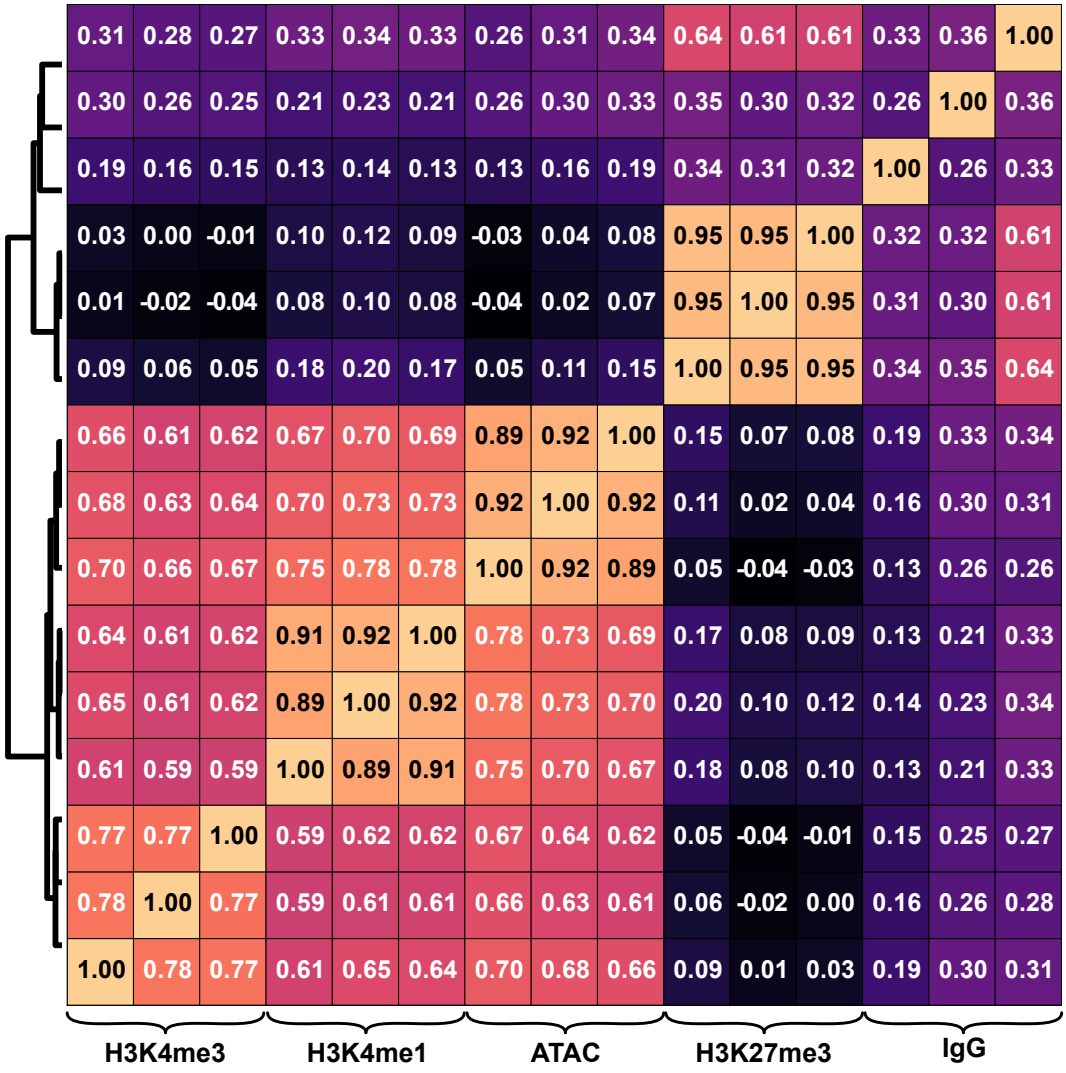


Figure B.10. Correlation analysis of genomic read distribution for *Hydra* ATAC-seq and CUT&Tag libraries shows reproducibility among biological replicates. Additionally, samples targeting active CREs (H3K4me1, H3K4me3, and ATAC-seq) were positively correlated with each other and showed no correlation to the repressive mark H3K27me3 or IgG controls.

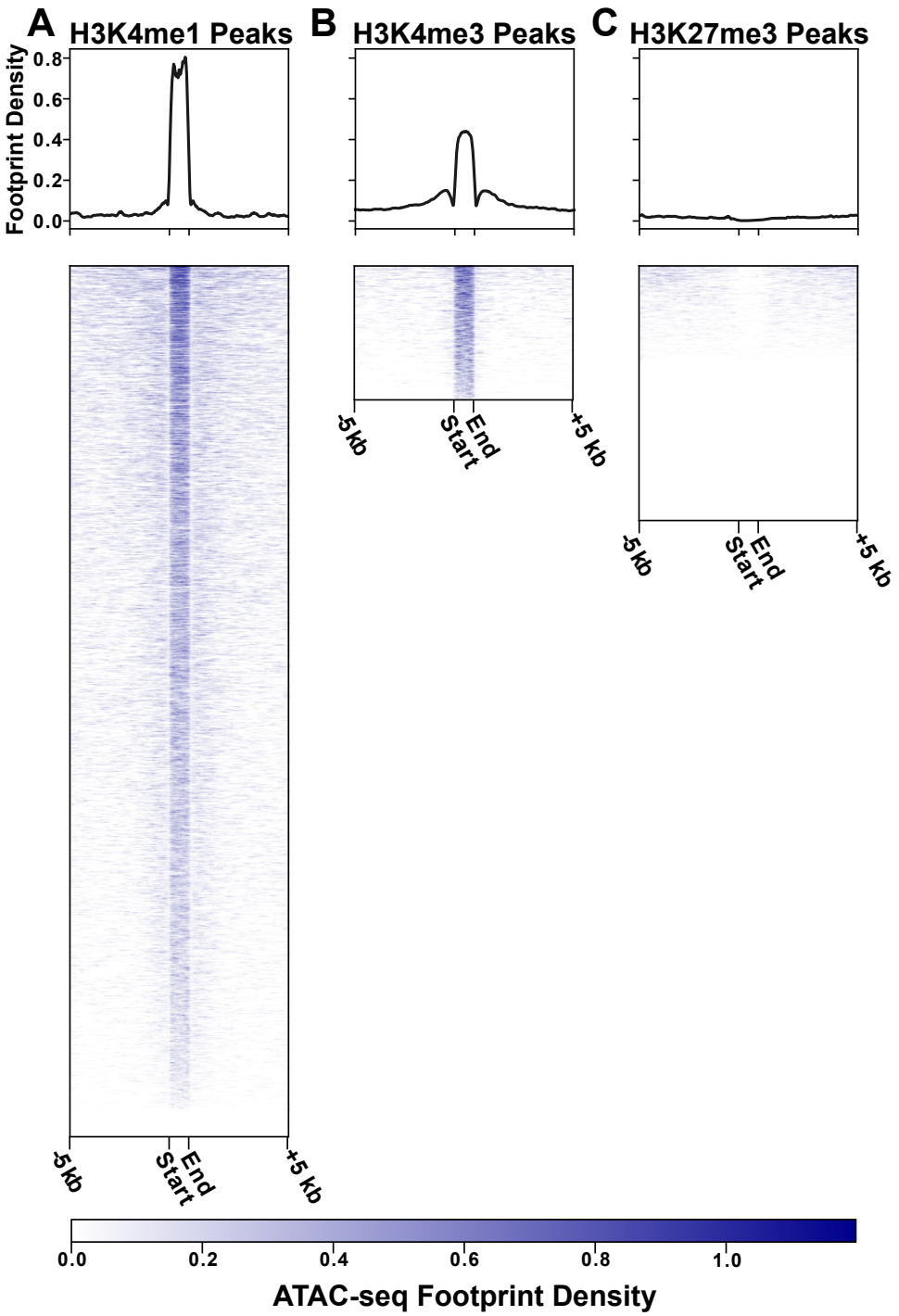


Figure B.11. Predicted transcription factor binding sites are enriched in regions with activating, but not repressive, histone marks. (A-C) ATAC-seq transcription factor binding footprints are enriched within (A) H3K4me1 and (B) H3K4me3 peaks, but not in (C) H3K27me3 peaks.

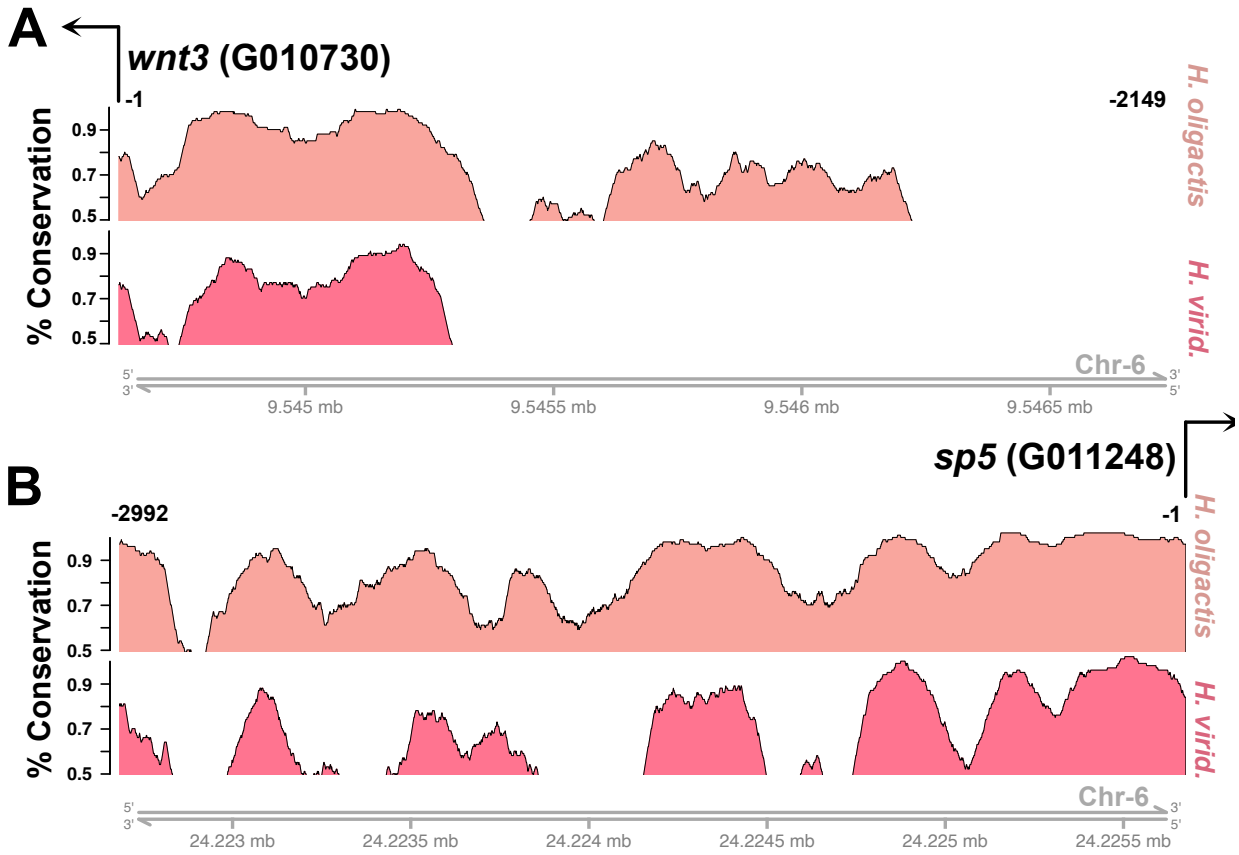


Figure B.12. Cross-species whole-genome alignments reveal conserved non-coding sequences in the strain AEP *H. vulgaris* genome. (A and B) 100 Bp moving window sequence conservation in sequence upstream of (A) *wnt3* and (B) *sp5* recapitulates previous results that used manual alignments (Vogg et al., 2019).

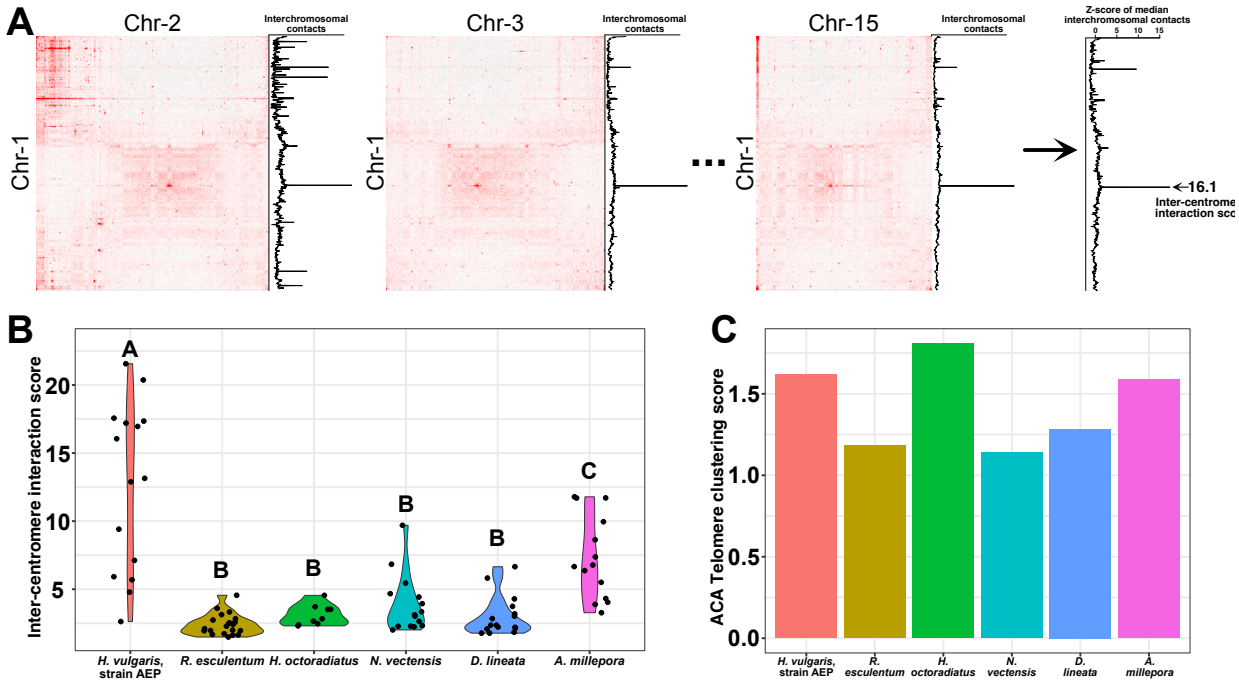


Figure B.13. The *Hydra* genome has significantly elevated rates of inter-centromeric, but not inter- telomeric interactions relative to other cnidarian genomes. (A) Visual summary of the approach used for unbiased quantification of inter-centromeric contacts. A z-score was calculated along the length of each chromosome based on the total number of inter-chromosomal contacts at each position. The inter-centromere interaction score was defined as the maximum z-score for each chromosome. Chromosomes with a strong inter-centromeric interaction signal will have high scores whereas chromosomes that lack such a signal with have low scores that do not rise far above the noise floor. (B) Distribution of inter-centromere interaction scores calculated for cnidarian genomes with available Hi-C data. Statistically distinct groups (Tukey's post-hoc test, $p \leq 0.05$) are labeled using letters above each species (i.e., species assigned the same letter are not statistically different, whereas species with different letters are). (C) Inter-telomeric interaction scores for cnidarians genomes with available Hi-C data calculated using a previously published methodology (Hoencamp et al., 2021).

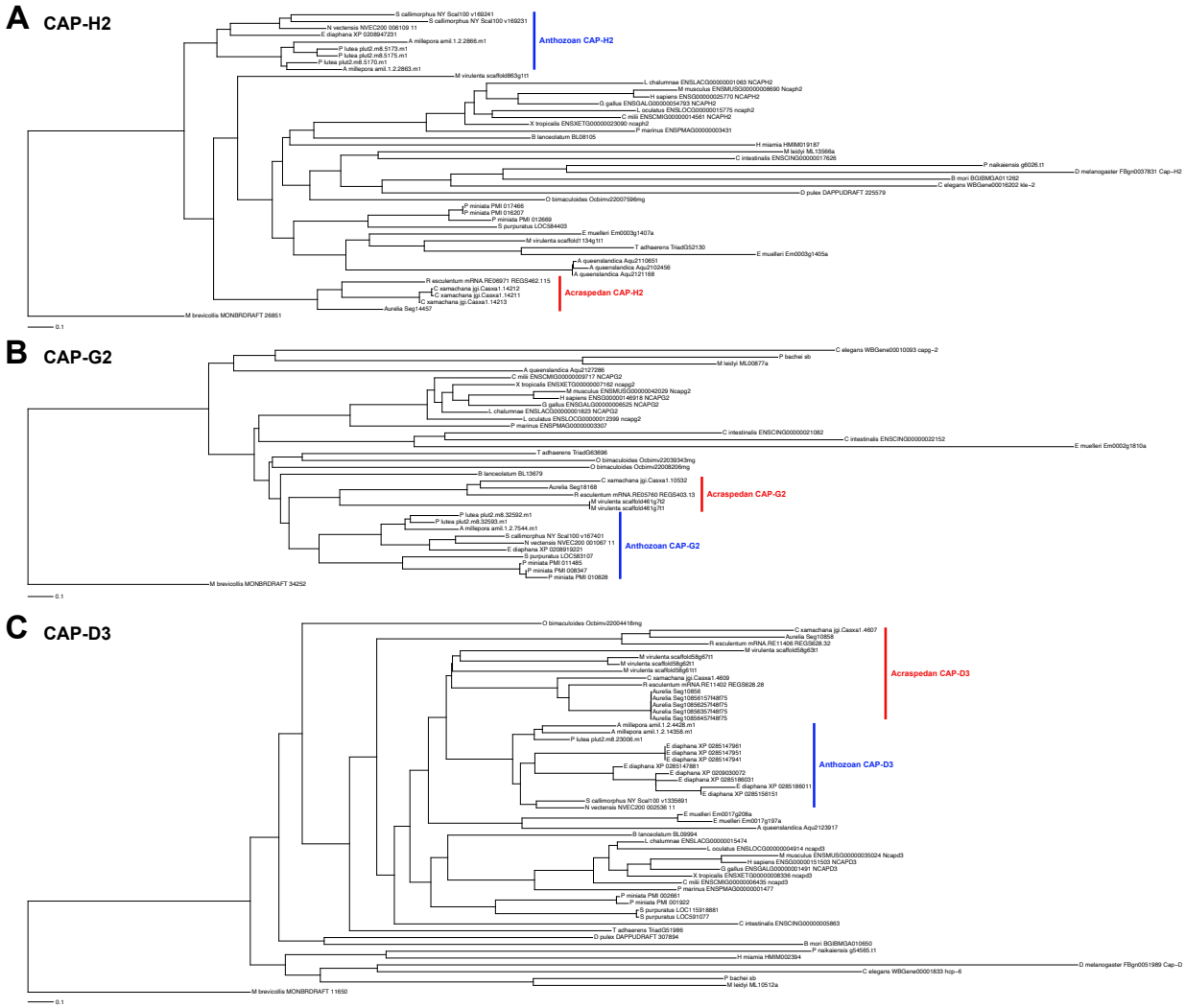


Figure B.14. Condensin II subunits are absent in hydrozoans. (A-C) Phylogenies constructed by Orthofinder identify orthologs of the condensin II subunits (A) CAP-H2, (B) CAP-G2, and (C) CAP-D3 in anthozoans and non-hydrozoan medusozoans (Acraspeda), but not in hydrozoans. The sequences used to construct these trees are provided in Supplemental Data S6.

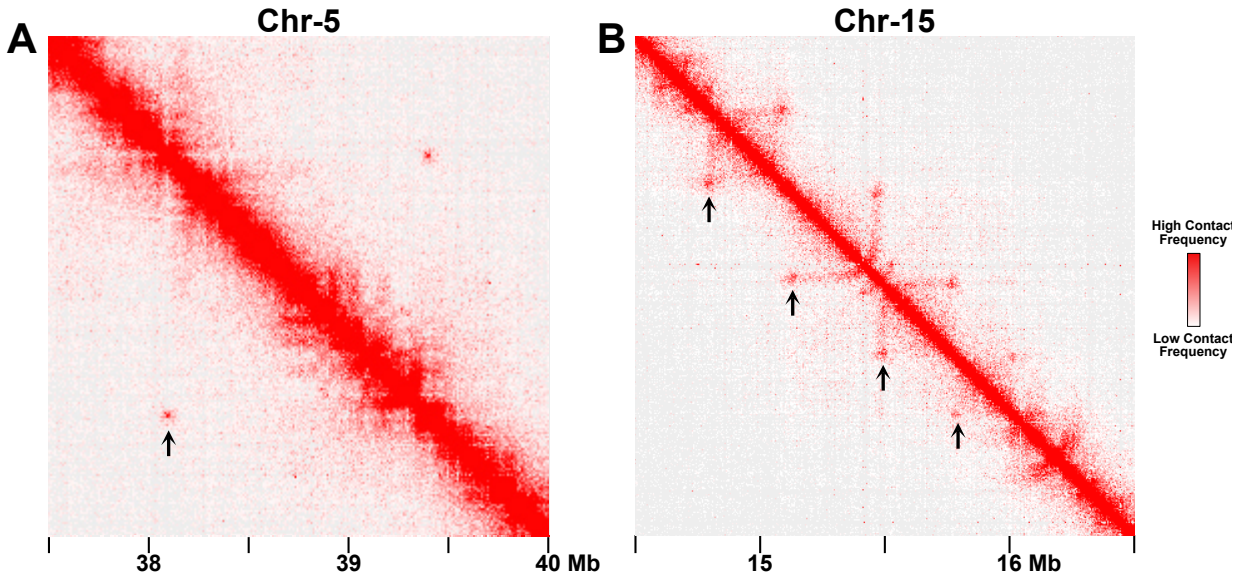


Figure B.15. Loop-like chromatin structures occur infrequently in the *Hydra* genome. Chromatin contact frequency maps for (A) Chr-5 and (B) Chr-15 include multiple off-diagonal dot-like interaction patterns, labeled with black arrows, that are distinctive of chromatin loops formed through the stable interaction of two distal loci.

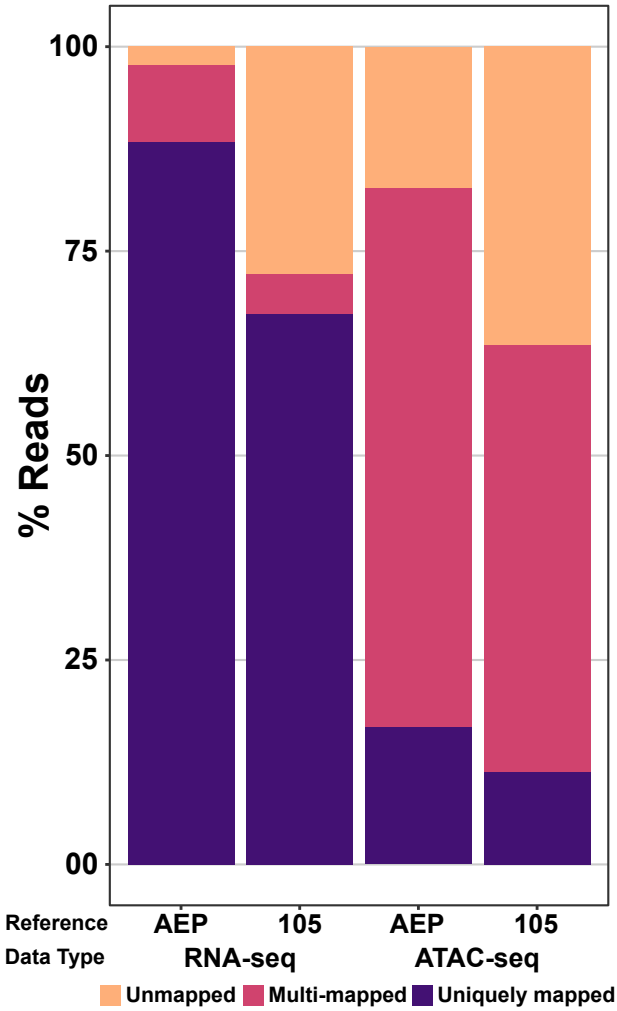


Figure B.16. Mapping efficiency of strain AEP *H. vulgaris* ATAC-seq and RNA-seq data are reduced when aligned to the strain 105 *H. vulgaris* genome reference.

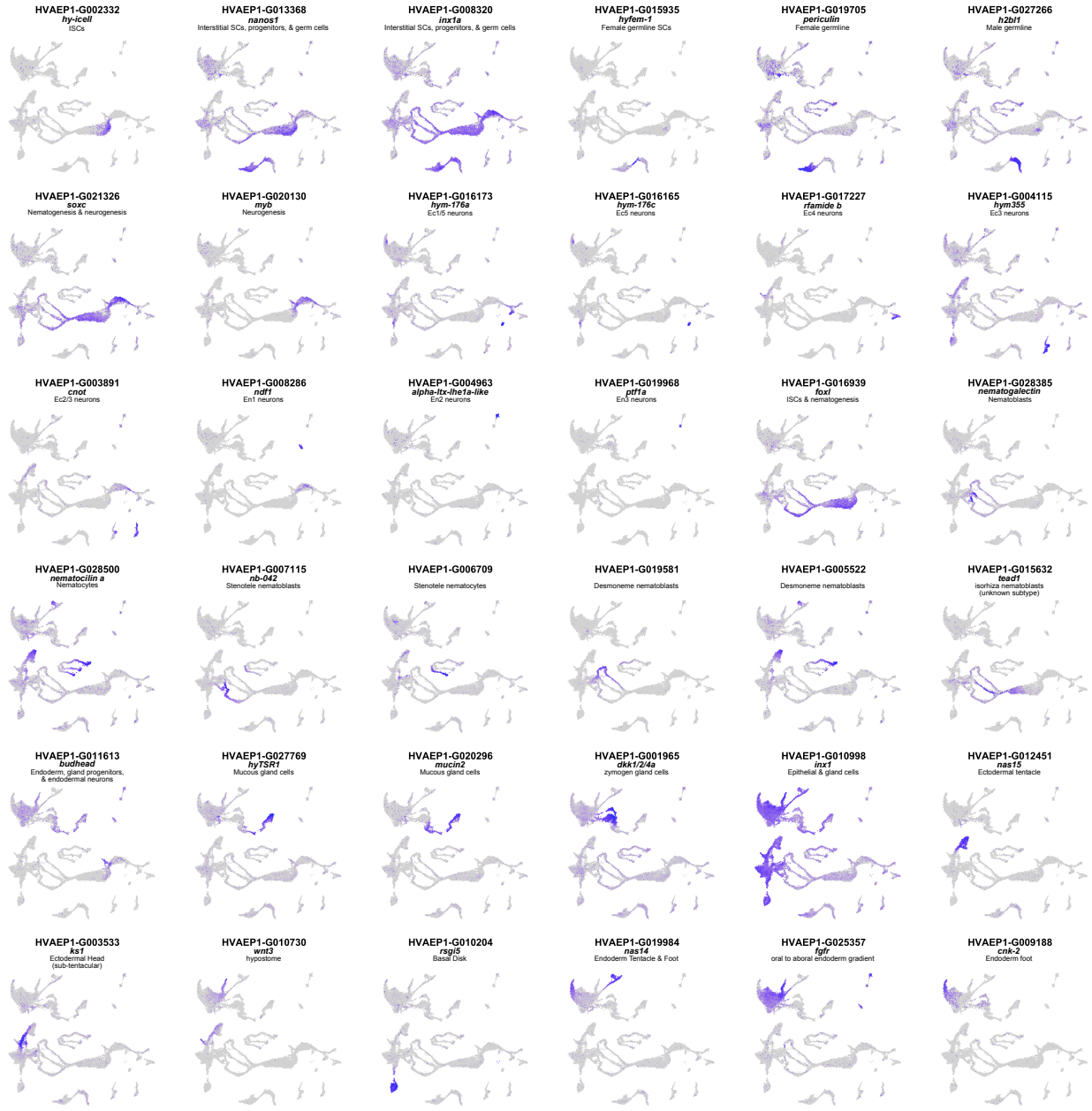


Figure B.17. Cluster annotation for the version of the strain AEP *H. vulgaris* single cell atlas that includes doublets using marker gene expression. All markers presented were validated in the initial atlas publication (Siebert et al., 2019).

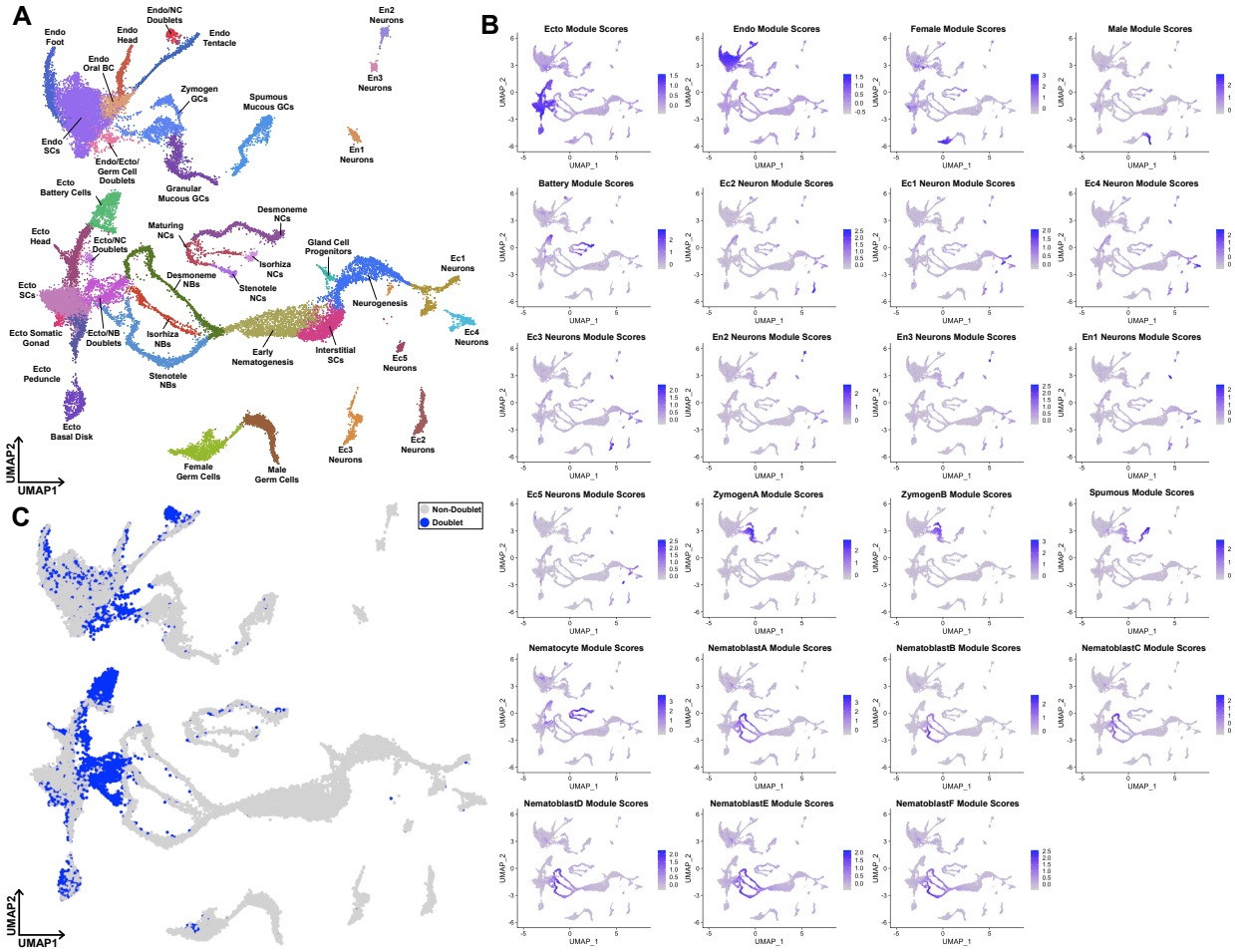


Figure B.18. Identification of doublets in the *Hydra* single-cell RNA-seq atlas. (A) Uniform Manifold Approximation and Projection (UMAP) dimensional reduction of the *Hydra* single-cell RNA-seq atlas mapped to the AEP reference genome prior to doublet removal. NCs, nematocytes; NBs, nematoblasts; SCs, stem cells; Ecto, ectodermal epithelial cells; Endo, endodermal epithelial cells; GCs, gland cells; Ec, neuron subtypes found in the ectoderm; En, neuron subtypes found in the endoderm. (B) Module scores for cell-type-specific gene expression programs as calculated by the Seurat AddModuleScore function. Cell-type-specific genes were identified using a Wilcoxon Rank Sum test as implemented in Seurat. (C) UMAP plot highlighting all cells identified as doublets. Doublets were defined as having a module score greater than 0.2 for both an epithelial module and any other cell type module.

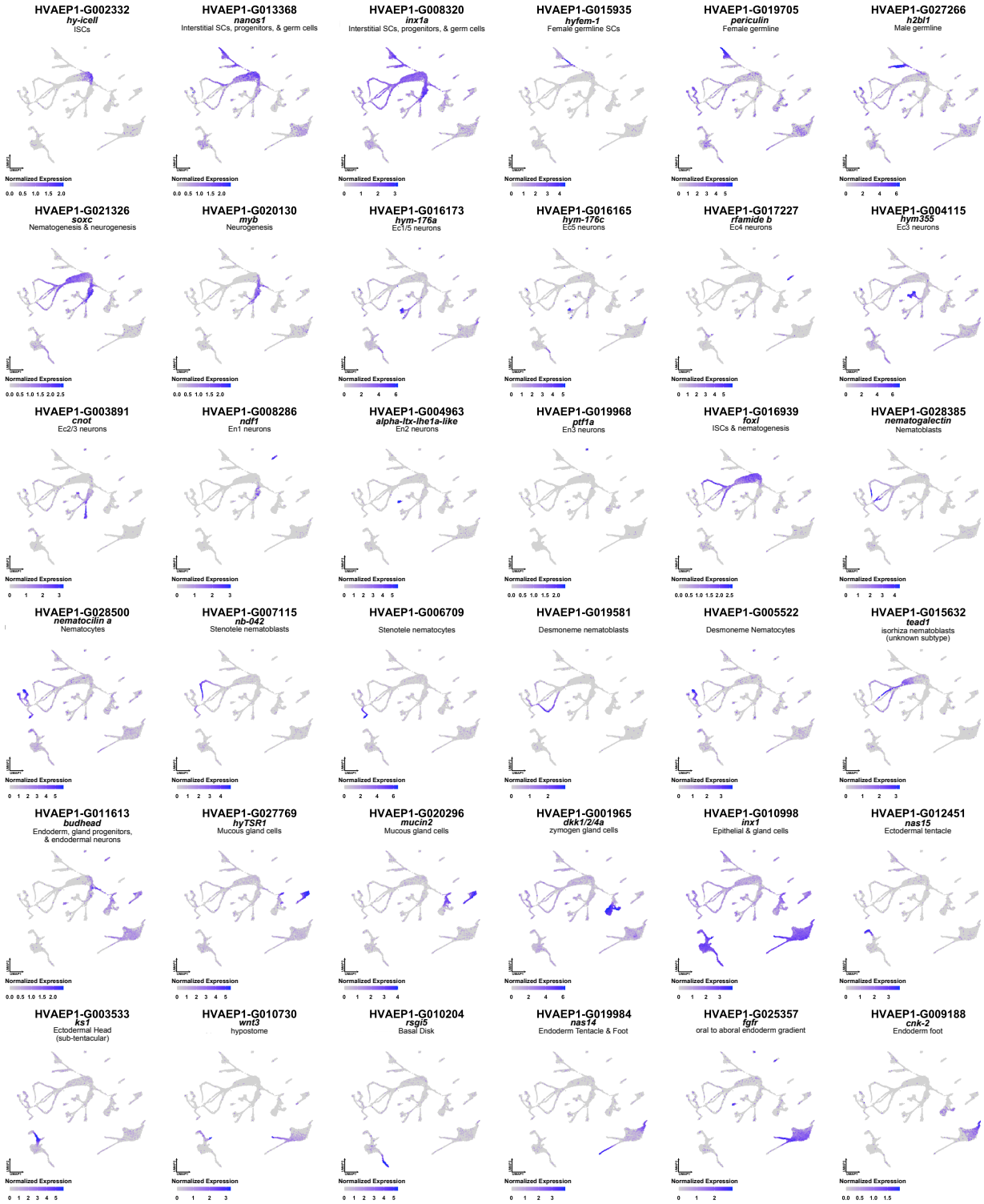


Figure B.19. Cluster annotation for the strain AEP *H. vulgaris* single cell atlas using marker gene expression. All markers presented were validated in the initial atlas publication (Siebert et al., 2019). The UMAP with labeled clusters is shown in Figure 4A.

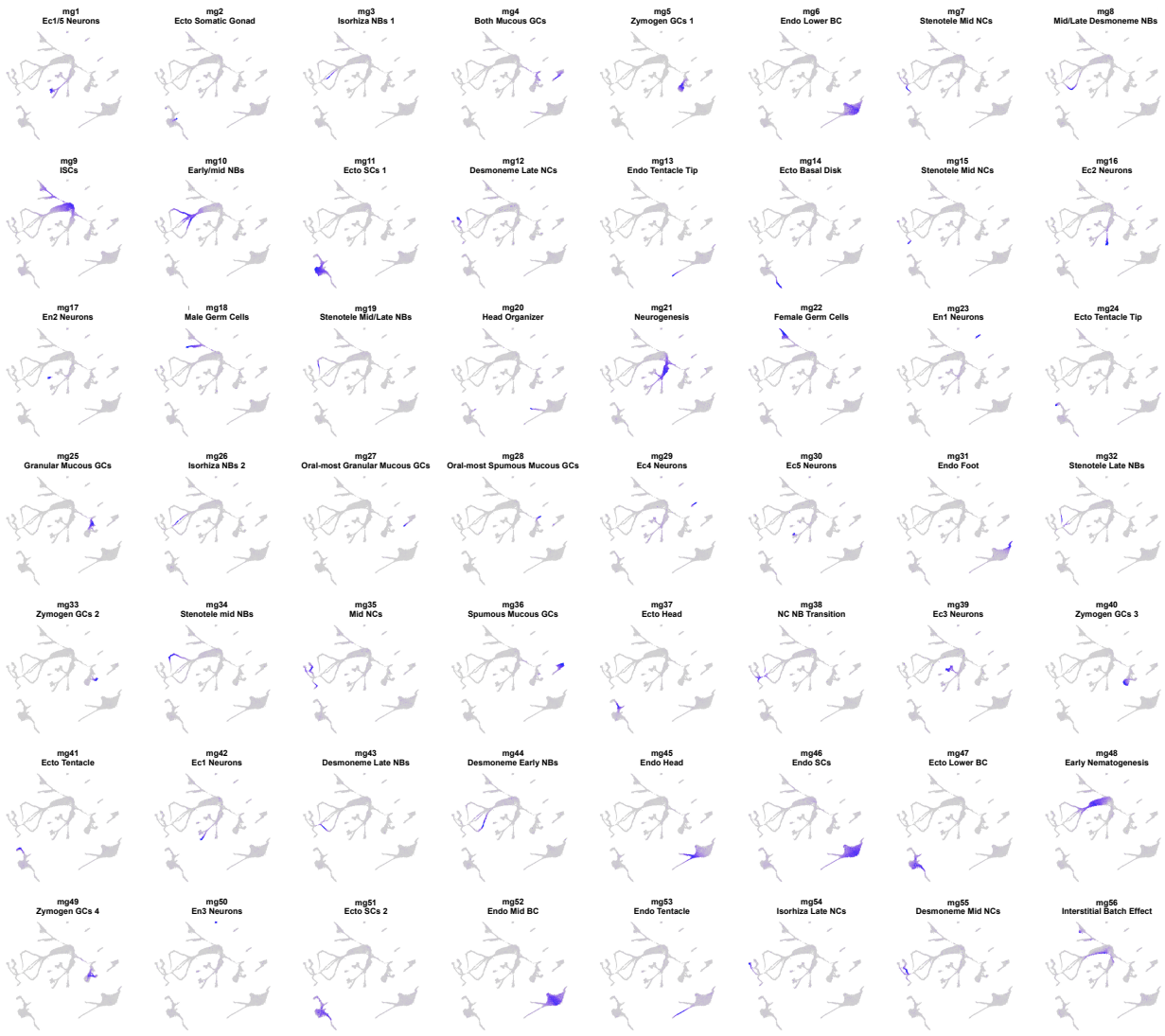


Figure B.20. Non-negative matrix factorization (NMF) identifies cell-type-specific co-expressed gene modules in the strain AEP *H. vulgaris* atlas. UMAP plots colored to highlight the cells expressing the 56 modules of co-expressed genes (i.e., metagenes) identified using NMF. More intense purple coloration indicates higher overall expression of a given metagene.

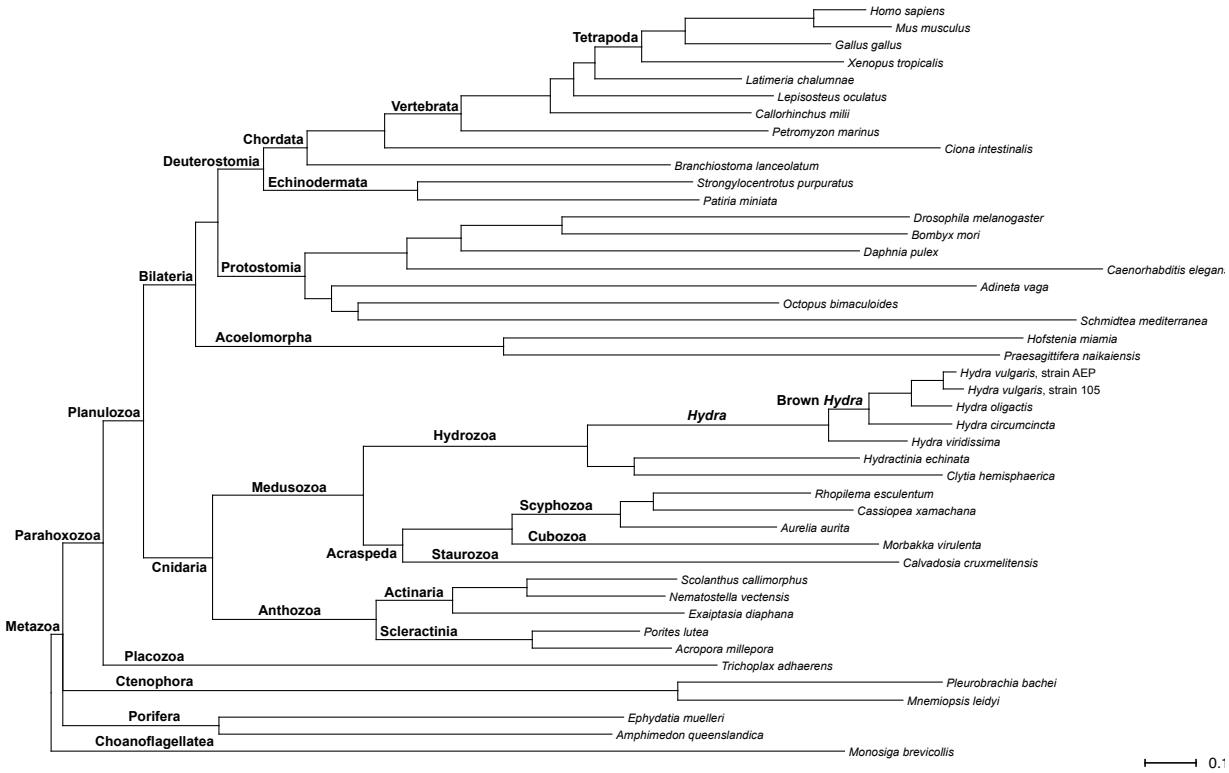


Figure B.21. Phylogeny of proteomes used in Orthofinder analysis. Proteome sources are provided in Table S5. Based on the tree branch lengths, the protein sequence divergence between *Hydra* and *Clytia* is roughly equivalent to that of humans and lampreys (*Petromyzon marinus*).

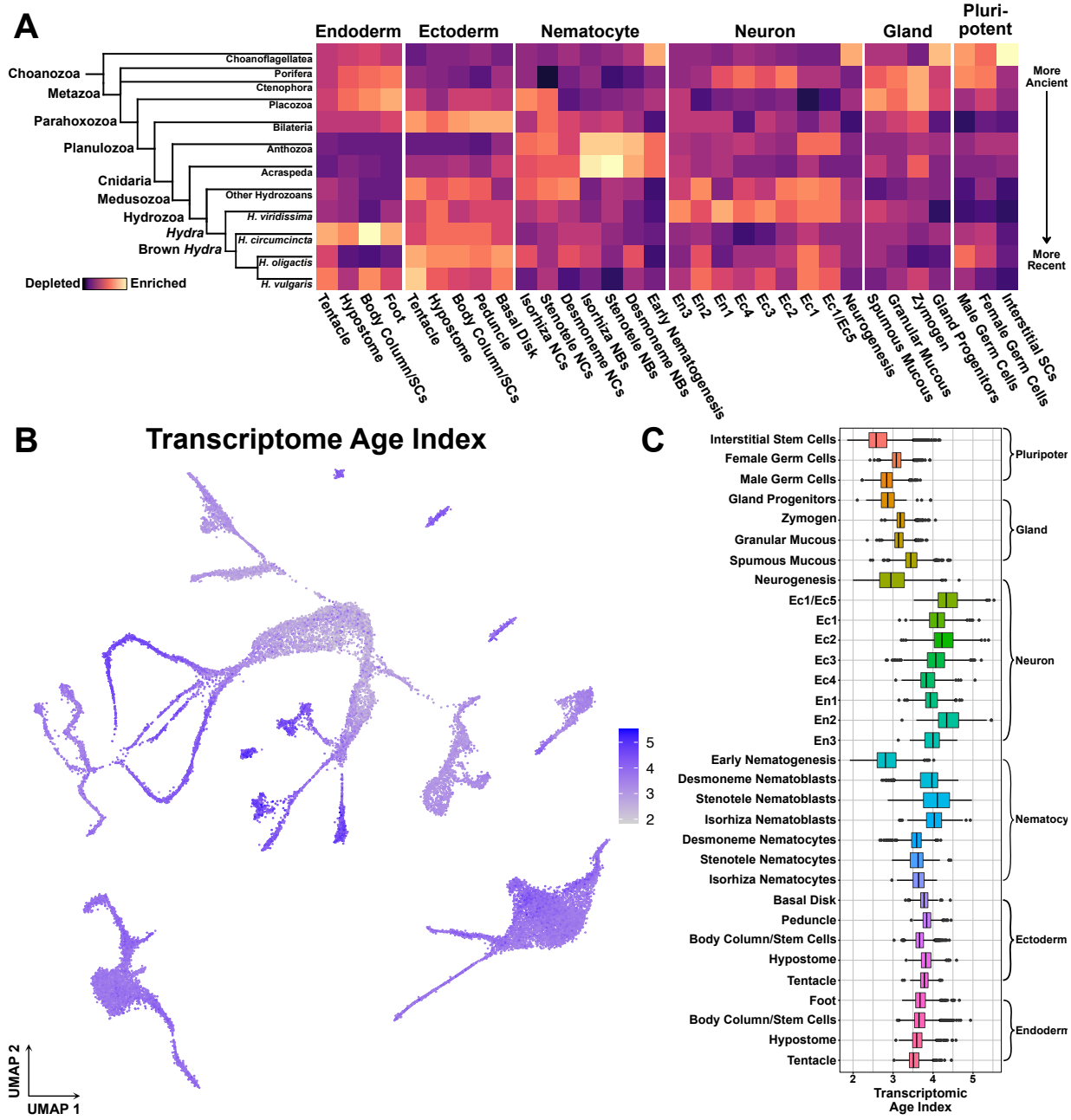


Figure B.22. Characterizing the relationship between gene age and cell-type-specific expression. (A) Heatmap depicting the relative enrichment of gene families by evolutionary age in the transcriptomes of different cell types suggest distinct evolutionary timelines. (B-C) Holistic quantification of single-cell transcriptome ages. The transcriptomic age index (TAI) is a weighted average that combines transcript abundance with gene age. High values of the resulting metric indicate a transcriptome is made up of relatively more recent genes and low values indicate a transcriptome is made up of relatively more ancient genes. (B) UMAP plot depicting TAI values for all single-cell transcriptomes in the *Hydra* cell atlas. (C) Boxplot of TAI values averaged by cell type.

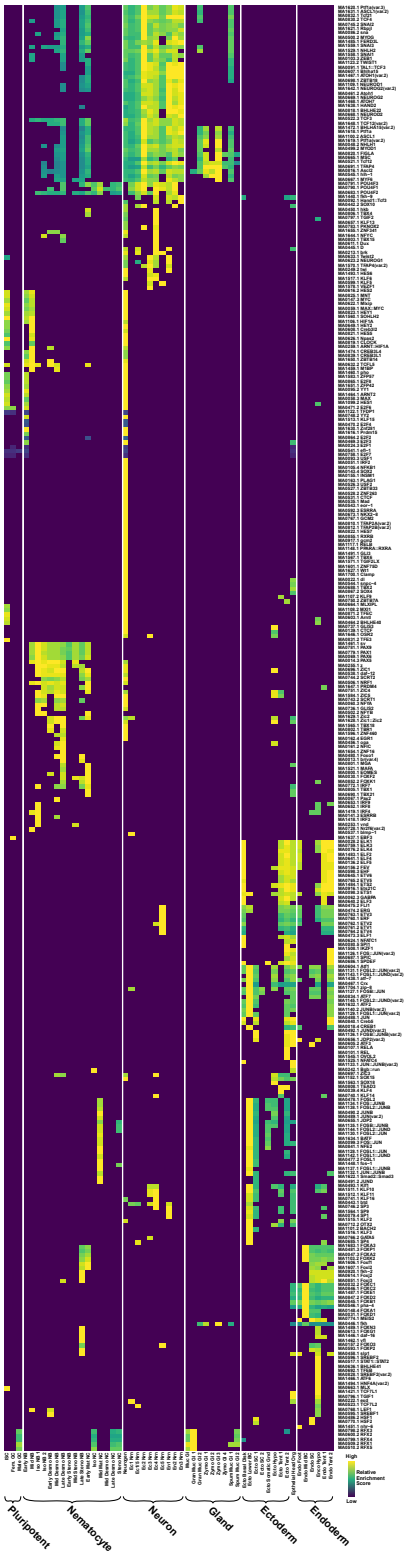


Figure B.23. Full motif enrichment results for the *Hydra* cell atlas. Enrichment scores that were not significant (adjusted P-value > 0.01) were set to zero. Heatmap values are normalized by row (i.e. by motif). Motifs are referred to using both their unique JASPAR ID (formatted as MA####.#) and the abbreviated name of their corresponding TF.

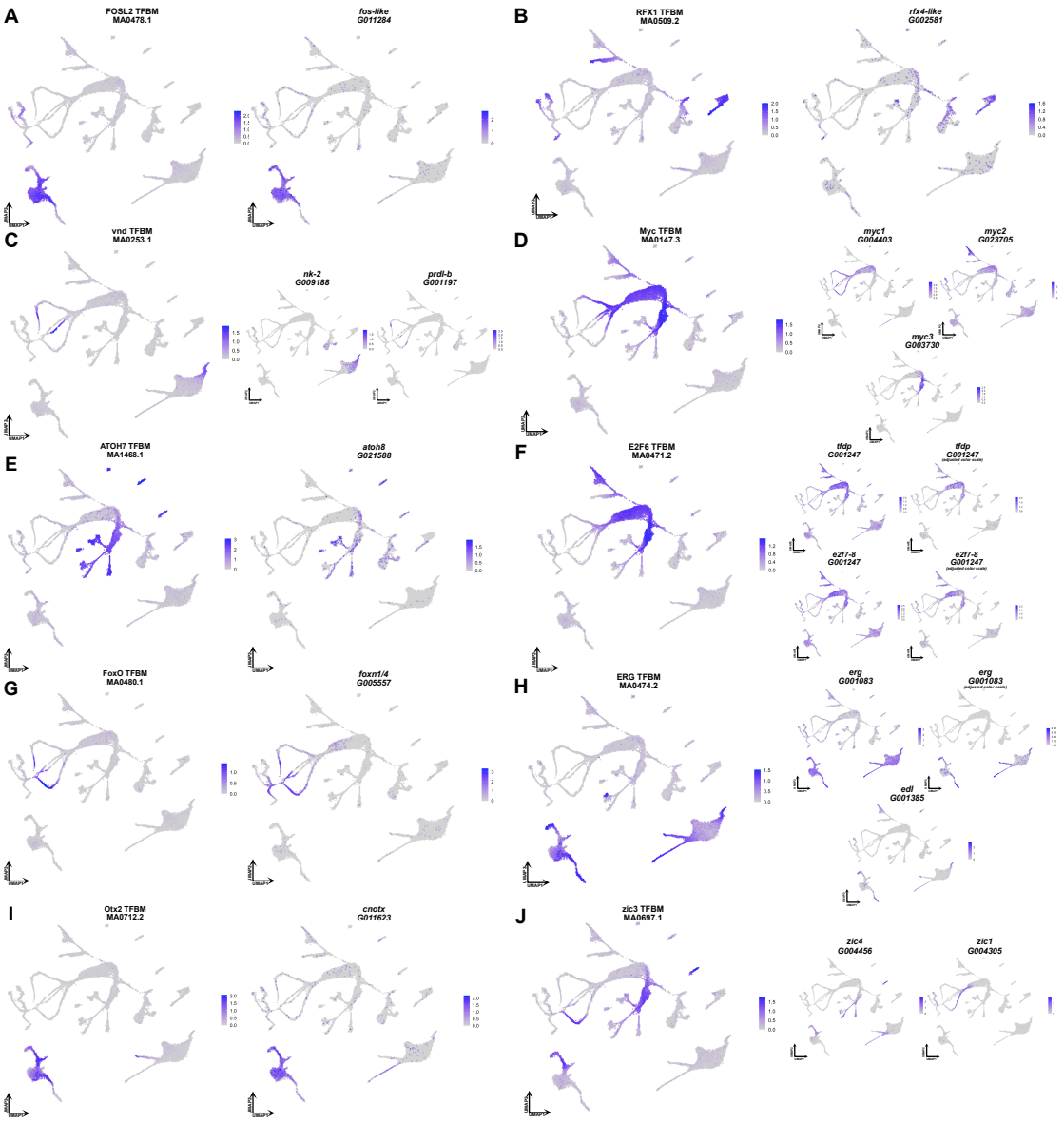


Figure B.24. Additional candidate regulators of gene co-expression in *Hydra*. Motif enrichment and gene expression correlation suggest that (A) *fos-like* is a regulator in ectodermal head and body column cells; (B) *rfx4-like* is a regulator in mucous gland cells; (C) the homeobox TFs *nk-2* and prdl-b are regulators in endodermal foot cells and nematoblasts respectively; (D) myc family transcription factors (TFs) are regulators in interstitial stem cells and progenitors; (E) *atoh8* is a regulator in mature and differentiating neurons; (F) e2f family TFs are regulators in interstitial stem cells, progenitors, and germ cells; (G) *foxn1/4* is a regulator in late nematoblasts; (H) ets family TFs are regulators in epithelial cells at the extremities (i.e., tentacle and foot tissue); (I) *cnotx* is a regulator in ectodermal cells in the body column, head, and tentacles; and (J) zic family TFs are regulators in ectodermal tentacle cells, Ec4 neurons, and desmoneme nematoblasts. Note that for some gene expression plots (*tfdb*, *e2f7-8*, and *erg*) two plots with different color scales are presented to highlight cells with high expression levels.

Color scales for motif plots refer to enrichment scores and normalized read counts in the gene expression plots.

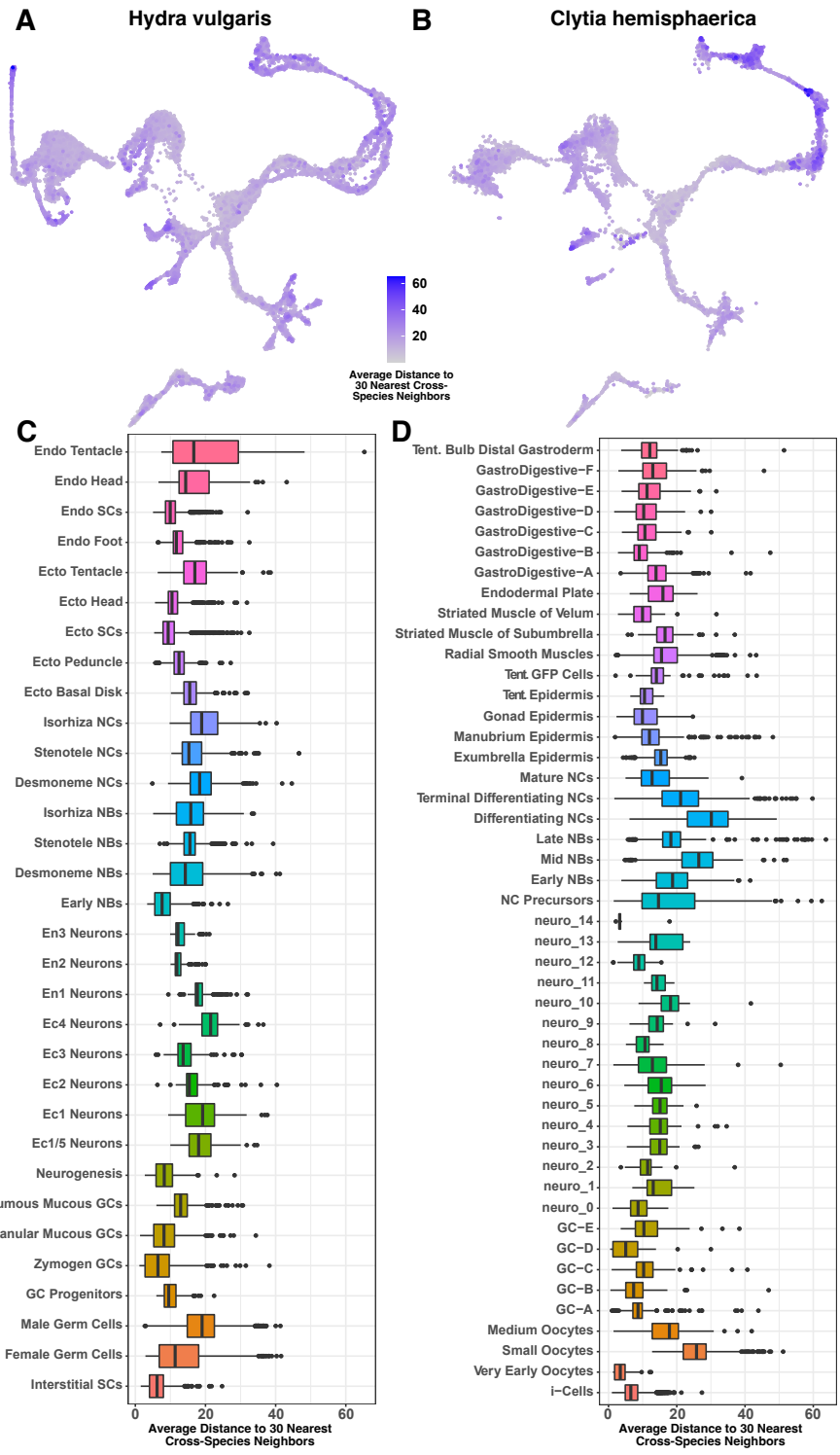


Figure B.25. Stem cells and early progenitors are generally associated with smaller inter-species alignment distances. Quantification of alignment distance in the cross-species *Hydra* and *Clytia* single-cell atlas. (A and B) UMAP plots depicting the average distance between (A) *Hydra* and (B) *Clytia* cells and their 30 nearest cross-species nearest neighbors in aligned

principal component space. Cells with lower distance values had transcriptional profiles that were more like cells from the other species. These values were calculated based only on one-to-one orthologs, and thus did not consider transcriptional differences based on genes unique to one of the species. (C and D) Box plots showing the distribution of distance scores for (C) *Hydra* and (D) *Clytia* grouped by cell type.

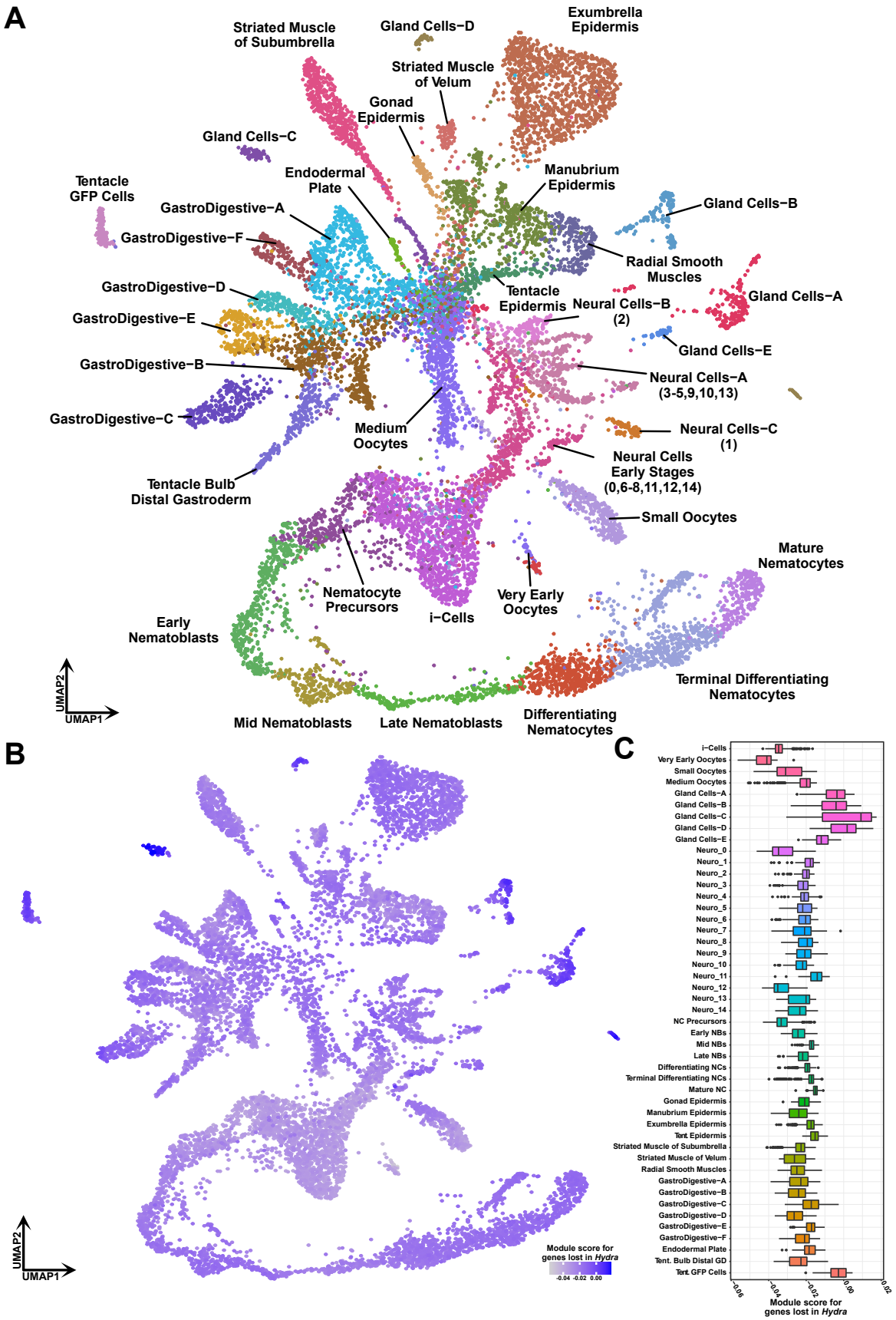

A
Striated Muscle of Subumbrella
Gland Cells−D
Exumbrella Epidermis
Striated Muscle of Velum
Gonad Epidermis
Gland Cells−C
Endodermal Plate
Manubrium Epidermis
Tentacle GFP Cells
GastroDigestive−A
Gland Cells−B
GastroDigestive−F
Radial Smooth Muscles
GastroDigestive−D
Tentacle Epidermis
Gland Cells−A
GastroDigestive−E
Neural Cells−B (2)
Gland Cells−E
GastroDigestive−B
Neural Cells−A (3-5,9,10,13)
GastroDigestive−C
Medium Oocytes
Neural Cells−C (1)
Neural Cells Early Stages (0,6-8,11,12,14)
Tentacle Bulb Distal Gastroderm
Small Oocytes
Mature Nematocytes
Nematocyte Precursors
Very Early Oocytes
i−Cells
Early Nematoblasts
Terminal Differentiating Nematocytes
UMAP2
UMAP1
Mid Nematoblasts
Late Nematoblasts
Differentiating Nematocytes
B
Module score for genes lost in Hydra
-0.04 -0.02 0.00
C
i−Cells
Very Early Oocytes
Small Oocytes
Medium Oocytes
Gland Cells−A
Gland Cells−B
Gland Cells−C
Gland Cells−D
Gland Cells−E
Neuro_0
Neuro_1
Neuro_2
Neuro_3
Neuro_4
Neuro_5
Neuro_6
Neuro_7
Neuro_8
Neuro_9
Neuro_10
Neuro_11
Neuro_12
Neuro_13
Neuro_14
NC Precursors
Early NBs
Mid NBs
Late NBs
Differentiating NCs
Terminal Differentiating NCs
Mature NC
Gonad Epidermis
Manubrium Epidermis
Exumbrella Epidermis
Tent Epidermis
Striated Muscle of Subumbrella
Striated Muscle of Velum
Radial Smooth Muscles
GastroDigestive−A
GastroDigestive−B
GastroDigestive−C
GastroDigestive−D
GastroDigestive−E
GastroDigestive−F
Endodermal Plate
Tent. Bulb Distal GD
Tent. GFP Cells
−0.06
−0.04
−0.02
0.00
0.02
Module score for genes lost in Hydra

Figure B.26. Transcripts expressed in *Clytia* gland and tentacle GFP cells are enriched for genes lost in *Hydra*. (A) Original annotated UMAP from the initial *Clytia* atlas publication (Chari et al., 2021). Parenthetical numbers under neuron cluster names refer to neuron subtypes contained within each broad neuron type. Neuron subtype names are based on a neural sub-clustering analysis from the initial atlas publication. Subtypes were assigned to the neuronal cluster that contained the largest portion of cells from a given subtype. (B-C) Module scores in the *Clytia* single-cell RNA-seq atlas were calculated based on a weighted average of the expression of all genes lost in *Hydra*. (B) UMAP plot depicting module scores for all single-cell transcriptomes in the *Clytia* atlas. (C) Module scores pooled by cell type.

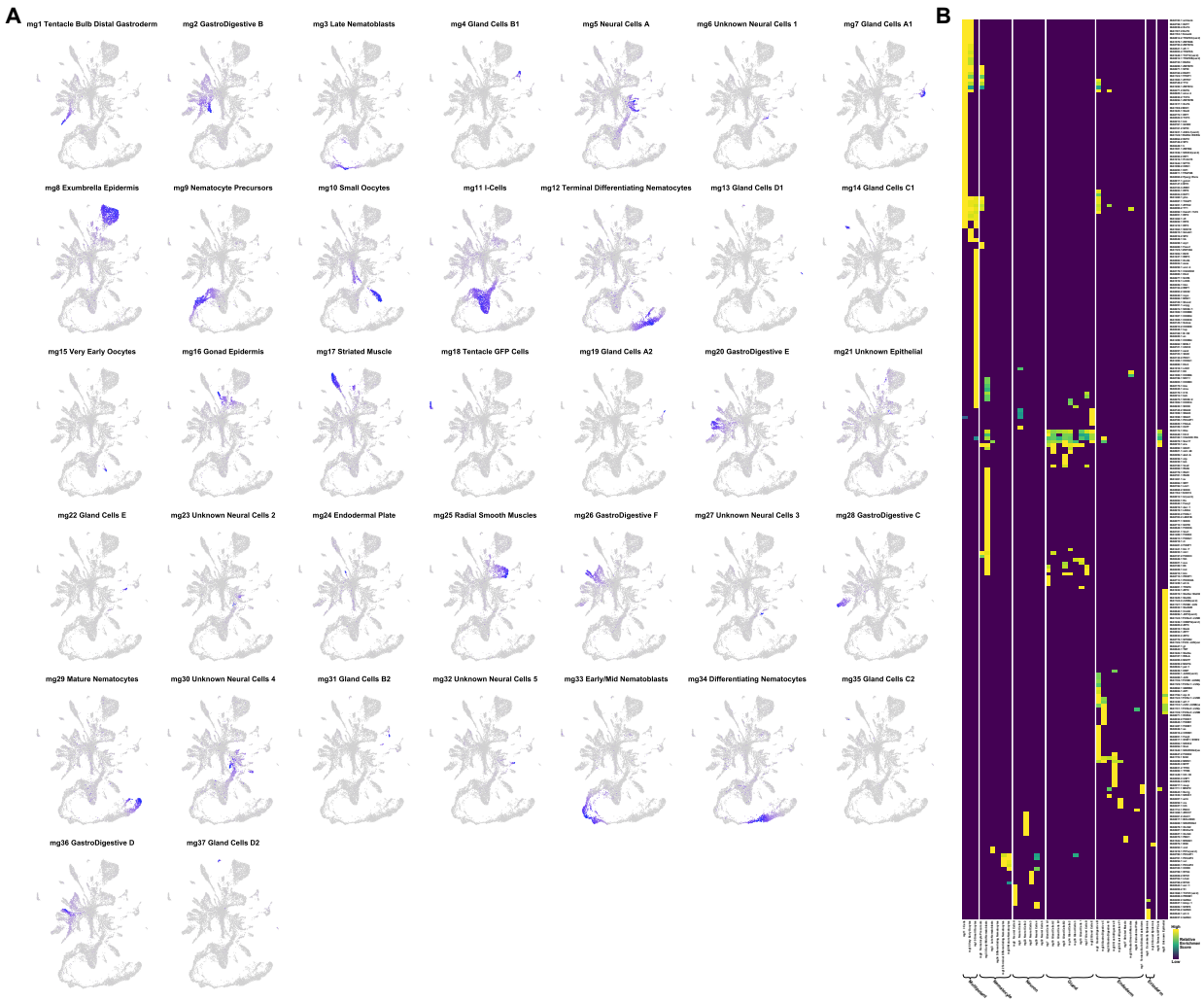


Figure B.27. Motif enrichment analysis in the *Clytia* single-cell medusa atlas. (A) UMAP plots from the original *Clytia* atlas publication (Chari et al. 2021) colored by non-negative matrix factorization (NMF) metagene expression. NMF identified 37 sets of co-expressed genes in the Clytia atlas, most of which could be readily assigned to previously annotated cell types. (B) Heatmap showing enrichment results for promoter proximal (≤ 1 kb from nearest TSS) sequences associated with the 37 metagenes identified by NMF. Sequences were assigned to metagenes based on gene weights generated as part of the standard NMF output. Values are presented only for enrichment results with an E-value < 10 (approximate adjusted p-value of 0.01). Motifs are referred to using both their unique JASPAR ID (formatted as MA####.#) and the abbreviated name of their corresponding TF.

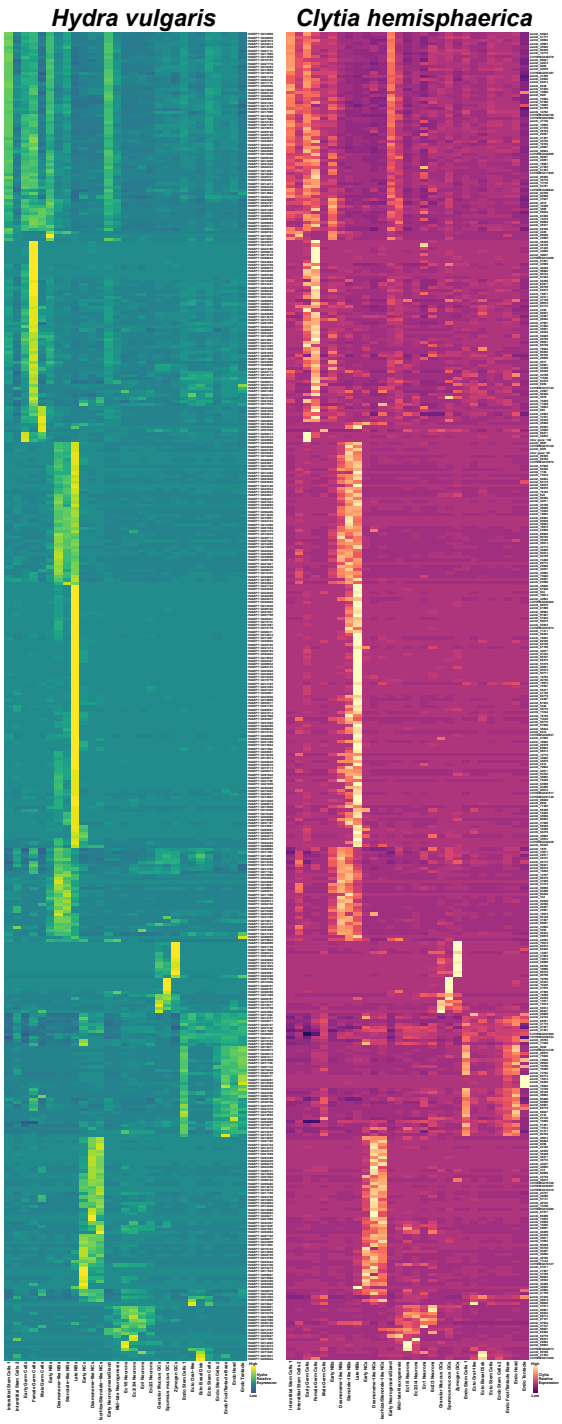


Figure B.28. Heatmap of orthologous gene pairs with similar cell-type-specific expression in *Hydra* and *Clytia* single-cell atlases. Gene pairs were classified as having similar expression patterns based on correlated expression (correlation score > 0.65) in the aligned cross-species principal component space. The clusters referred to in the heatmap column names refer to a fine-resolution cross-species Louvain clustering analysis presented in Figure B.29.

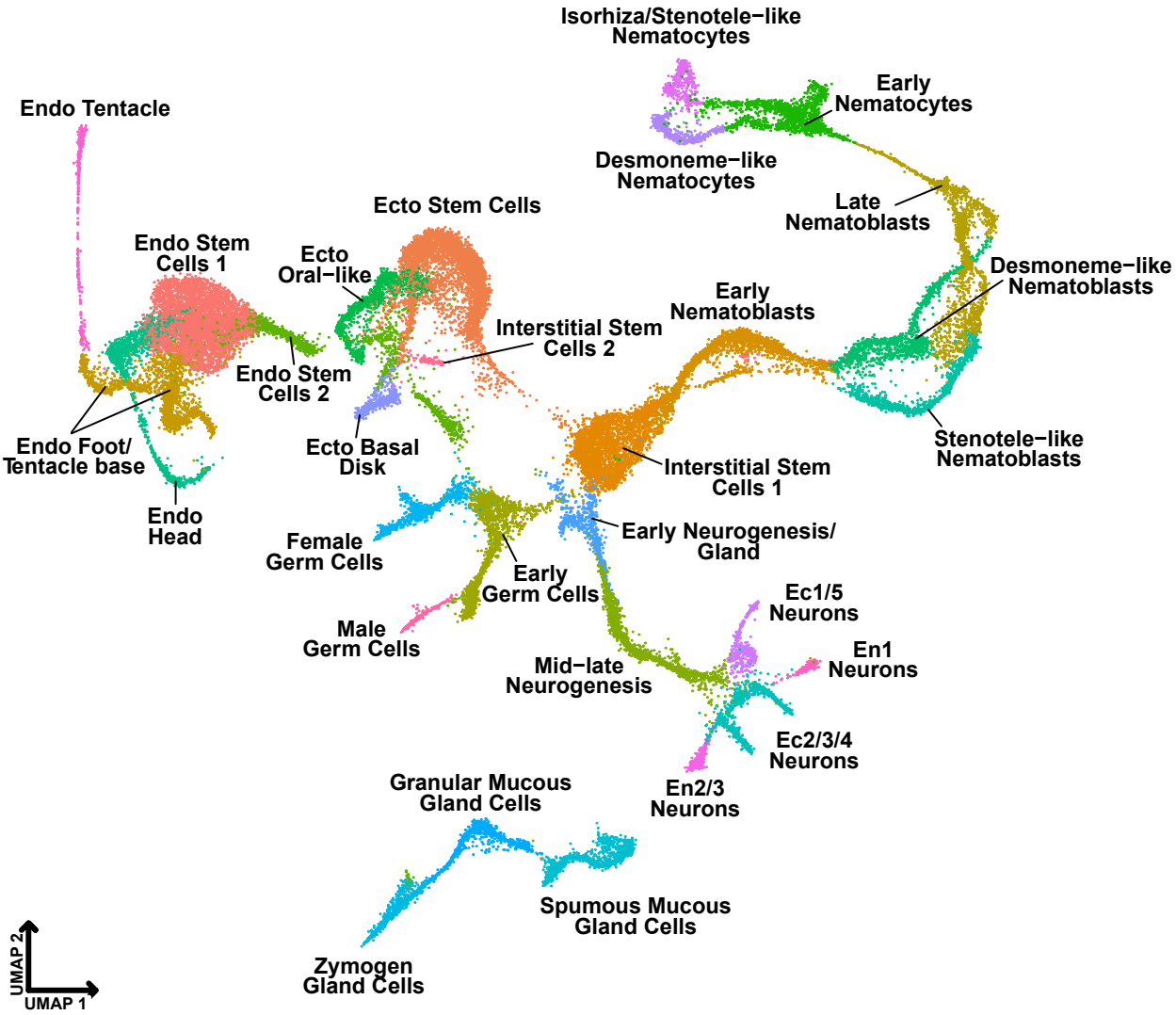


Figure B.29. Cross-species aligned *Clytia* and *Hydra* UMAP colored by the Louvain clusters used for the expression correlation heatmaps in Figures B.28 and B.30.

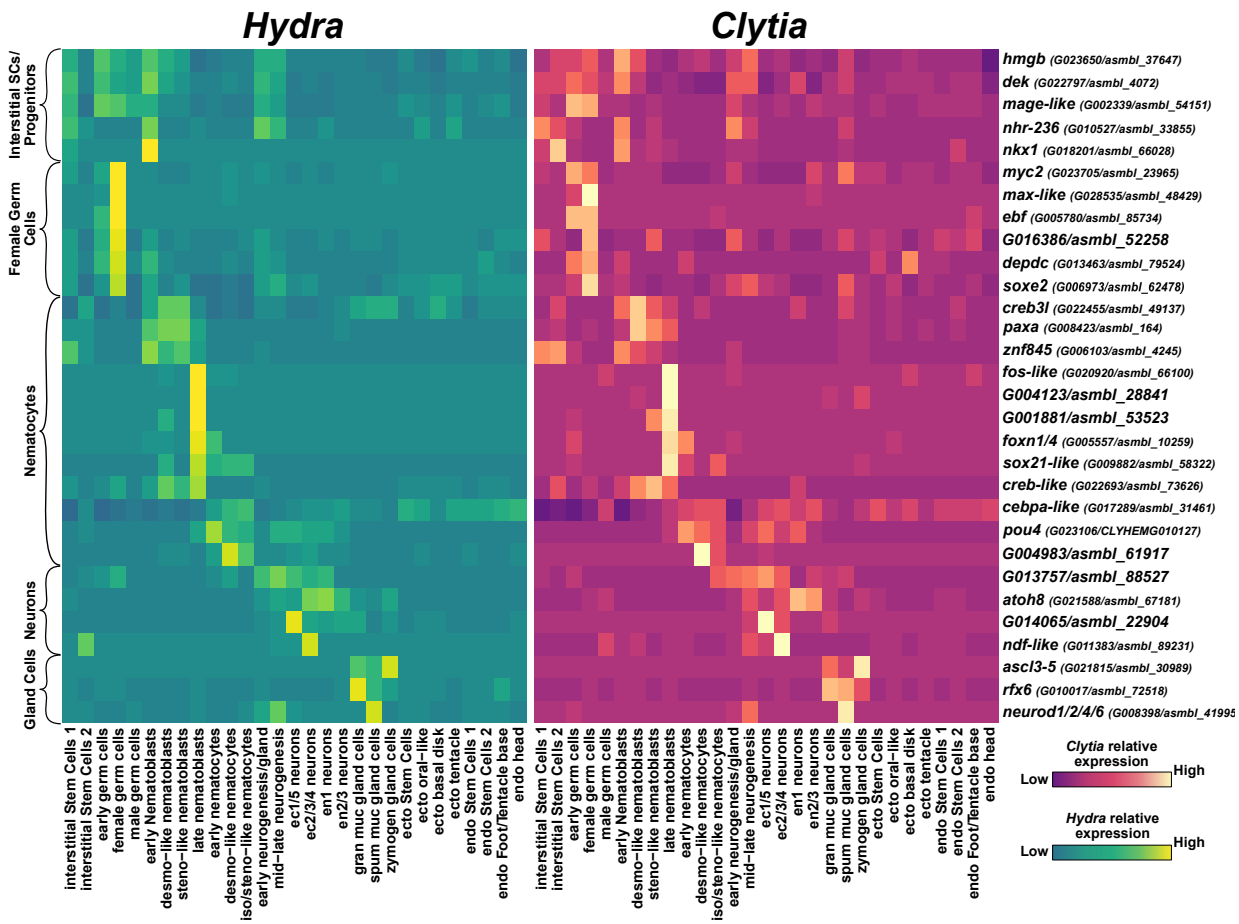


Figure B.30. Heatmap of predicted transcription factors (TFs) with similar cell-type-specificity in *Hydra* and *Clytia*. TFs were predicted based on the presence of a predicted DNA-binding domain. Orthologous gene pairs were classified as having similar expression patterns based on correlated expression (correlation score > 0.65) in the aligned cross-species principal component space. The heatmap column names refer to a fine-resolution cross-species Louvain clustering analysis presented in Figure B.29. Heatmap values are normalized by row.

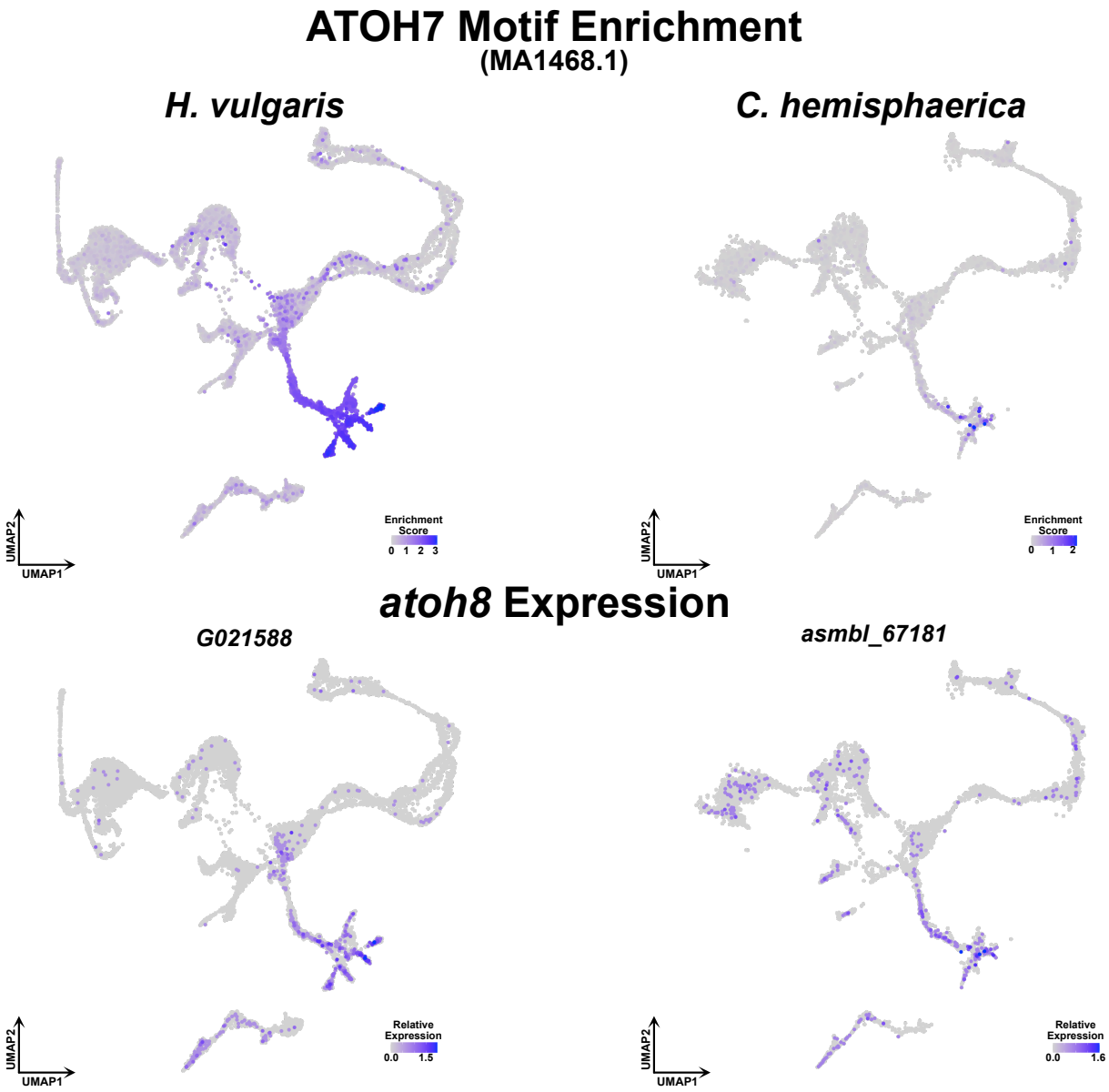


Figure B.31. Motif enrichment and gene expression patterns in the *Hydra* and *Clytia* cell atlases suggest *atoh8* is a conserved regulator of hydrozoan neurogenesis.

Supplemental tables and data can be found at DOI: 10.1101/gr.277040.122

Appendix C: Multiple nerve rings coordinate *Hydra* mechanosensory behavior

This chapter was originally published in *eLife*:

Badhiwala KN, AS Primack, CE Juliano, and JT Robinson. 2021. Multiple nerve rings coordinate Hydra mechanosensory behavior. eLife *10:e64108.*

I made the following contributions to the work presented in appendix C: I generated the transgenic line *Tg(ef1a:GCaMP7b)*$^{cj1\text{-}en}$ used for calcium imaging in this study. I also performed qPCR to validate the location of neuron subtypes in the *Hydra*, wrote methods, and provided feedback on the manuscript. I specifically contributed to figures C.1, C.3, and C.15.

ABSTRACT

Hydra vulgaris is an emerging model organism for neuroscience due to its small size, transparency, genetic tractability, and regenerative nervous system; however, fundamental properties of its sensorimotor behaviors remain unknown. Here, we use microfluidic devices combined with fluorescent calcium imaging and surgical resectioning to study how the diffuse nervous system coordinates *Hydra*'s mechanosensory response. Mechanical stimuli cause animals to contract, and we find this response relies on at least two distinct networks of neurons in the oral and aboral regions of the animal. Different activity patterns arise in these networks depending on whether the animal is contracting spontaneously or contracting in response to mechanical stimulation. Together, these findings improve our understanding of how *Hydra*'s diffuse nervous system coordinates sensorimotor behaviors. These insights help reveal how sensory information is processed in an animal with a diffuse, radially symmetric neural architecture unlike the dense, bilaterally symmetric nervous systems found in most model organisms.

INTRODUCTION

Discovering the fundamental principles of neural activity and behaviors requires studying the nervous systems of diverse organisms. Animals have evolved different neural structures like the nerve net (e.g., *Hydra*), nerve cords and ganglia (e.g., *Caenorhabditis elegans*, *Aplysia*, planaria), and brain (e.g., *Drosophila*, zebrafish, rodents, and primates). Despite the vastly different structures, many behaviors are conserved across species, including sensorimotor responses (Ahrens et al., 2012; Chen et al., 2018; Clark et al., 2013; Ghosh et al., 2017; Haesemeyer et al., 2018; Kaplan et al., 2018) and sleep (Artiushin and Sehgal, 2017; Gandhi et al., 2015; Guo et al., 2018; Hill et al., 2014; Kanaya et al., 2020; Kayser and Biron, 2016; Raizen et al., 2008; Zhdanova et al., 2001; Zimmerman et al., 2008). By comparing neural circuits that support similar behaviors despite different architectures, we can discover organizational principles of neural circuits that reflect millions of years of evolutionary pressure. While there are many potential organisms that would support this type of comparative neuroscience, only a small group of animals have the qualities to support laboratory experiments: short generation span, ease of breeding and manipulation in laboratory conditions, small and compact size, optical transparency, and a well-developed genetic toolkit with a complete spatial and molecular map of the nervous system.

Transparent, millimeter-sized animals in particular offer a number of advantages for neuroscientists because it is possible to image neural activity throughout the entire nervous system using genetically encoded calcium or voltage-sensitive fluorescent proteins (Ahrens et al., 2013; Broussard et al., 2014; Chen et al., 2013; Cong et al., 2017; Gonzales et al., 2020; Kim

et al., 2017; Lemon et al., 2015; Portugues et al., 2014; Prevedel et al., 2014; Vladimirov et al., 2014). In addition, some millimeter-sized animals are compatible with microfluidic devices for precise environmental control and microscopy techniques that offer cellular-resolution functional imaging of the entire nervous system. These properties, combined with genetic tractability, provide a powerful way of revealing neuronal dynamics across the entire nervous system (not just a small region) during behaviors. For instance, whole nervous system imaging of confined or freely moving animals has revealed the neuronal dynamics underlying locomotion (Nguyen et al., 2016; Prevedel et al., 2014) and sensory-motivated global state transitions in *C. elegans* (Gonzales et al., 2019; Nichols et al., 2017), responses to noxious odor and visuomotor behaviors in zebrafish (Ahrens et al., 2013; Cong et al., 2017; Kim et al., 2017; Portugues et al., 2014; Prevedel et al., 2014; Vladimirov et al., 2014), responses to light and odor in *Drosophila* (Aimon et al., 2019; Lemon et al., 2015), and neuronal ensembles correlated with basal behaviors and response to light and heat in *Hydra* (Badhiwala et al., 2018; Dupre and Yuste, 2017).

Hydra is unique among the small, transparent organisms discussed above due to its regenerative ability and highly dynamic nervous system. While most small, transparent model systems (like *C. elegans* or zebrafish larvae) suffer permanent behavioral deficits from the loss of one or a few neurons (Bargmann and Avery, 1995; Bargmann and Horvitz, 1991; Bejjani and Hammarlund, 2012; Hecker et al., 2020; Kroehne et al., 2011), *Hydra* can completely recover from a significant neuronal loss to regain normal contractile behavior in as little as ~48 hr (Gierer et al., 1972; Itayama and Sawada, 1995; Soriano et al., 2009). This radially symmetric freshwater cnidarian has a nervous system composed of two diffuse networks of neurons, one embedded in

the endoderm and another embedded in the ectoderm (Burnett and Diehl, 1964; Lentz and Barrnett, 1965). While *Hydra's* diffuse nerve net is highly dynamic with continuous cellular turnover and migration (Bode et al., 1988; Richard D. Campbell, 1967), regions with increased neuron density resembling nerve rings have a comparatively lower neuronal turnover (Figure C.1 A) (Bode et al., 1973; Epp and Tardent, 1978; Hufnagel and Kass-Simon, 2016; Koizumi et al., 1992). One of these regions is in the oral end in the apex above the ring of tentacles ('hypostomal nerve ring'), and another is in the aboral end in the foot ('peduncle nerve ring') (Figure C.1 A). Recent single-cell RNA sequencing has provided a complete molecular and spatial map of the *Hydra* nervous system, including identification of unique cell-type-specific biomarkers to generate new transgenic models (Siebert et al., 2019). This existing molecular and spatial map of the nervous system suggests that there is no overlap in the neuronal cell types that make up the hypostomal and peduncle nerve rings, and the distribution of the cell types varies along the length of the body (Figure C.1 B). Finally, the demonstration of microfluidic and transgenic tools combined with *Hydra's* dynamic yet 'simple' neural architecture has enabled observations of basal and sensory motivated behaviors in the regenerating nervous system (Badhiwala et al., 2018).

To better establish *Hydra* as a model organism for comparative neuroscience, it is critical to understand their basic sensorimotor behaviors, such as response to touch. While it is well documented that *Hydra* contract when mechanically agitated or poked with a pipette (Mast, 1903; Rushforth, 1965; Rushforth et al., 1963; Rushforth and Burke, 1971; Wagner, 1905), we found no quantitative reports of how this behavior depends on stimulus intensity or is mediated by neural activity. Although significant insights in *Hydra* behavior have been made over the last

several decades with simple methodologies and manual observations, including that the tentacles and/or the hypostome are needed for mechanosensory response, these experiments lack quantitative characterization of neuronal or behavioral response. Forceps-induced touch allows local stimulation, but it is difficult to control the force applied manually (Campbell, 1976; Takaku et al., 2014). While stimulation with mechanical agitation allows control over stimulus intensity, these observations are limited to changes in body lengths (Rushforth et al., 1963).

Here, we use whole-animal functional imaging combined with resection studies to discover that despite the apparently diffuse nerve net in *Hydra*, these animals process sensorimotor responses in specialized regional networks. To study the mechanosensory response, we first developed a microfluidic system to apply a local mechanical stimulus and quantify *Hydra's* behavioral and neural response. We then measured these responses in the absence of select regions of the body and found at least one of the neuron-rich regions, the hypostome (oral) or the peduncle (aboral), is required to coordinate spontaneous contractions, though the oral network plays a more significant role. We found a significant reduction in the mechanosensory response with the removal of the hypostome, the region where sensory information is likely processed. These sensorimotor experiments combined with whole-animal neural and epitheliomuscular imaging reveal that *Hydra* is capable of receiving sensory information along the body column; however, the oral region is necessary for coordinating the motor response.

RESULTS

Hydra's mechanosensory response is dependent on stimulus intensity

To better understand sensory information processing in *Hydra*, we developed a double-layer microfluidic system that can apply a local mechanical stimulus while we image the response of the entire nervous system using fluorescence microscopy (Figure C.2, Video C.1). This local mechanical stimulation is made possible by push-down microfluidic valves (Figure C.2 A) that deliver mechanical stimuli to a portion of the *Hydra* body with precise temporal and spatial control (see Materials and methods). For all experiments, we pressurized a valve (400 μm diameter) that was directly above the animal (for 1 s every 31 s, see Materials and methods) to stimulate the body column while simultaneously performing functional calcium imaging (Figure C.2 B). We selected the middle of the body for stimulation region to help ensure that we stimulated roughly the same region of the animal throughout each experiment. This choice was based on the observation that the body column region was relatively stationary, whereas the oral and aboral extremities had large displacements during body contractions and elongations.

Experiments showed that this stimulation paradigm delivered a local mechanical stimulation with most of the mechanical force localized to a radius of approximately 250 μm around the microfluidic valve. To measure the locality of this stimulus, we performed an experiment using transgenic *Hydra* (nGreen) (Siebert et al., 2019) (expressing GFP pan-neuronally (and in neural progenitors) and tracked the position and fluorescence intensity (GFP) from individual neurons during mechanical stimulation (N = 222 neurons over 1 min). When the microfluidic valve was pressurized to deliver mechanical stimulation, we found significantly increased average cellular (or tissue) displacement ($p<0.001$). Further analysis of the cellular movements showed the spatial

distribution of mechanical force from the stimulation was primarily experienced by the neurons directly under the valve (Figure C.5). Specifically, we found that the tissue directly below the valve was compressed (z direction) when mechanically stimulated – the neurons directly under the valve had a small magnitude of lateral (x–y direction) displacement. The tissue bordering the valve was stretched away from the valve center – the neurons in the neighboring regions around the valve had the largest lateral displacement. This lateral displacement decreased for neurons that were farther from the center of the valve. Neurons more than 750 μm from the microfluidic valve center showed a negligible displacement of less than 5 μm (95% CI lower bound = 5.8 μm), which is ~550% less than the displacement of neurons bordering the valve.

Having established our method to provide local mechanical stimuli, we characterized *Hydra's* sensitivity to local touch and the associated neural response. We performed experiments using transgenic *Hydra* expressing GCaMP6s in neurons (Dupre and Yuste, 2017). When we delivered mechanical stimuli, we found bright calcium signals generated by a small number of neurons in the hypostome and body column and a striking co-activation of many neurons in the ectodermal peduncle nerve ring (Figure C.2 B, Videos C.2 and C.3). This nerve ring activity appeared as either a single bright calcium spike (or 'contraction pulse') or a volley of bright calcium spikes (or 'contraction burst'). We also found that the calcium-sensitive fluorescence averaged over a region of interest (ROI) surrounding the peduncle faithfully represented the contraction pulses and bursts measured from individual neurons (Figure C.6). When we analyzed single-neuron calcium dynamics from this peduncle nerve ring, we found extremely high correlated activity as previously reported for contractions pulses and bursts (Figure C.6) (Badhiwala et al., 2018; Dupre and Yuste, 2017).

Given the similarity of these data between large and small ROIs, we chose to use the peduncle ROI to measure neuronal contraction bursts and pulses because it does not require single-neuron tracking, which significantly increased the throughput of our data analysis. We further confirmed that this signal is not the result of motion artifacts by measuring fluorescence from *Hydra* (nGreen) that express GFP pan-neuronally using a similar ROI. In that case, we did not see the strong fluorescence signals associated with contraction pulses and bursts (Figure C.7 E-F). Body length proved to be an unreliable quantification of contractions due to the stimulation artifacts (Figure C.7 D); however, we were able to accurately measure muscle activity associated with contractions by imaging calcium spikes in the epithelial muscle cells (Figure C.8 and C.9 and Videos C.4 and C.5). Based on these experiments, we define *Hydra's* 'mechanosensory response' as calcium spikes in neural activity from the peduncle ROI and the associated calcium spikes in the epithelial muscles from the whole body if they occur within 1 s of mechanical stimulation onset (Figures C.2 C-D and C.10).

Using the neuronal fluorescence calcium imaging described above, we found that the probability of the mechanosensory response depends on the intensity of the stimulus, which is consistent with many psychometric functions (Figure C.2 E). *Hydra* were five times more likely to contract within 1 s of receiving a strong mechanical stimulus than during a random 1 s interval without a stimulus (stimulus valve pressure 20 and 25 psi; response probability = 0.60 ± 0.06 and 0.77 ± 0.08, mean ± SEM, respectively; no stimulus valve pressure = 0 psi; response probability = 0.11 ± 0.01, mean ± SEM; Figure C.2 E and H). During mild stimuli, there was a slight increase (~2×) in response probability above the spontaneous activity, although this increase was not

statistically significant compared to spontaneous contraction bursts or pulses (valve pressure 10 and 15 psi; response probability = 0.23 ± 0.01 and 0.24 ± 0.03, mean ± SEM, respectively; Figure C.2 E and H). We found that *Hydra* did not respond to a weak mechanical stimulus that corresponded to a valve pressure of 5 psi (response probability = 0.09 ± 0.02, mean ± SEM; Figure C.2 E and H).

Further analysis of the calcium activity pattern revealed that the single contraction pulses (calcium spikes in the peduncle neurons) were more frequent when we repeated mechanical stimulation every 31 s for 1 hr. While spontaneous contraction pulses or bursts were observed roughly once every minute in microfluidics, when we stimulated *Hydra* with a strong mechanical stimuli, we found that the frequency nearly matched the 31 s between stimuli (Interval between spontaneous contraction bursts or pulses = 72.98 s ± 4.58, mean ± SEM; 66.68 ± 5.73 s median ± SE; Interval between stimulated contraction bursts or pulses = 20 psi, 38.28 ± 1.77 s, mean ± SEM; 31.14 ± 2.21 s, median ± SE; 25 psi, 32.20 ± 1.04 s, mean ± SEM; 31.14 ± 1.30 s, median ± SE; Figure C.2 F and H). While the majority of spontaneous calcium spikes formed bursts, stimulated calcium spikes were roughly three times more likely to be a single contraction pulses (percentage of spontaneous spiking activity that is a single contraction pulse = 0 psi, 16.87 ± 5.38% mean ± SEM; percentage of stimulated spiking activity that is a single contraction pulse = 20 psi, 56.84 ± 2.48%; 25 psi, 62.50 ± 5.23%, mean ± SEM; Figure 2g, h).

Hydra sensitivity to mechanical stimuli is lowest near the aboral end

Because of the diffuse neural architecture of *Hydra*, we expected each patch of *Hydra* tissue to be equally responsive to mechanical stimuli. However, when we stimulated transgenic *Hydra*

expressing GCaMP6s (N = 3 whole animals, stimulated 40 times per region with 22 psi) at three different regions along their body (oral, middle body, and aboral, stimulated 40 times at each region with 22 psi), we found the aboral end of *Hydra* to be less sensitive than the center of the body (aboral region response probability = 0.1 ± 0.025; mid-body region response probability = 0.42 ± 0.025; mean ± SEM; $p<0.01$; Figure C.11 A). Epitheliomuscular calcium imaging (N = 8 whole animals expressing GCaMP7b in endodermal epitheliomuscular cells, stimulated 40 times in body column region with 22 psi) also showed that the sensitivity to mechanical stimulation generally decreases towards the aboral end of the *Hydra* (Figure C.11 B). Furthermore, we found that the difference in sensitivity along the body was not an artifact due to differently sized *Hydra* experiencing different pressures from the microfluidic valves. We observed no statistically significant trend between animal size and response probability (Figure C.11 C). These findings, combined with the transcriptional analysis and in situ hybridizations that indicate higher density of sensory neurons in the oral half of the *Hydra* (Siebert et al., 2019), suggest that the oral end may be more sensitive to mechanical stimuli. Unlike other organisms that have unique motor responses like reversals or acceleration depending on the location of mechanical stimuli (e.g., *C. elegans*) (Chalfie et al., 1985; Wicks et al., 1996), we observed that the same motor program was initiated regardless of where on the body the animal was touched. The only difference we observed was the fact that the response probability depended on where along the oral and aboral axis we delivered the mechanical stimulus.

Hydra's mechanosensory response is mediated by electrically coupled cells

Based on the latency of the mechanosensory response, we hypothesized that sensorimotor information is transmitted by electrical activity in the *Hydra* and not by passive calcium

diffusion through epithelial cells. This hypothesis is supported by the fact that aneural *Hydra,* which no longer spontaneously contract in the absence of stimuli, are capable of aversive contractile response to touch; however, their responses are slow and require strong mechanical stimulus (Campbell et al., 1976; Takaku et al., 2014). To further test our hypothesis, we created a transgenic *Hydra* strain that expresses the calcium indicator GCaMP7b (under the EF1α promoter, see Materials and methods) in the endodermal epitheliomuscular cells. During body contractions, both endodermal and ectodermal epitheliomuscular cells are co-activated (Wang et al., 2020). With this transgenic line we measured contraction pulses and contraction bursts by averaging calcium activity in all epitheliomuscular cells (Figures C.7 and C.9). We hypothesized that if the mechanosensory response was primarily mediated by calcium diffusion through epitheliomuscular cells, we would expect to see propagation of calcium activity from the site of the stimulation. However, this was not the case; we observed fast propagation of calcium activity throughout the entire endoderm. Imaging calcium activity in the peduncle neurons showed that changes in neural activity correlated with the observed changes in epithelial muscle cells with increasing stimulus intensity (Figures C.2 and C.9).

This suggests that the mechanosensory response is indeed mediated by the electrically coupled cells. Although *Hydra*'s behavioral responses are slow compared to other invertebrates, the 0.5 s average response time in our data could not be explained by calcium diffusion alone, which would take ~100 s to travel the average distance of ~0.5–1 mm between the stimulation site and the peduncle or hypostome (Figure C.12 A). Our data show that the calcium signals from the peduncle neurons or endodermal muscles start increasing within ~0.1–0.2 s following stimulation onset, reaching peak fluorescence at ~0.5–0.6 s regardless of the stimulus intensity

(Figure C.12 B-E). We found no difference in the calcium response times between neural and muscle calcium imaging, though this could be influenced by the dynamics of the calcium indicator, which typically cannot give information about latencies less than 50 ms (Chen et al., 2013).

Aboral neurons are not necessary for mechanosensory response in *Hydra*

Having quantitatively established *Hydra's* sensorimotor response to mechanical stimuli, we next asked if specific regions of the nervous system play primary roles in mediating this response. Patch-clamp electrophysiology of individual neurons in *Hydra* has thus far been unsuccessful despite attempts by many research groups, and minimally invasive, cell-type specific neuromodulation techniques such as optogenetics have yet to be developed for *Hydra*. However, the animal's regenerative abilities allow us to resect large portions of tissue without killing the animal, thus we can borrow from the tradition of lesioning brain regions to study their functions (Krug et al., 2015; Passano and McCullough, 1964; Pierobon, 2015; Rushforth et al., 1963; Vaidya et al., 2019).

Because the *Hydra* body plan is radially symmetric with cell types primarily varying along the oral-aboral axis of the body column (Figure C.1), we chose to make axial cuts across the body column to remove select neuronal populations from the animal. Our rationale was that these resections would remove entire or nearly entire groups of neuronal cell types. We then allowed ~6–12 hr for the animal to recover. This recovery time helps to reduce the confounding contributions from initial tissue regeneration and allows animals to recover enough to tolerate microfluidic immobilization. This period is long enough to allow the wounds to close, the

molecular response to injury to be completed (Cazet and Juliano, 2020; Tursch et al., 2020), and the initial molecular events of regeneration to start; but it is *not* long enough for the animal to regenerate lost neurons, which takes approximately 30–72 hr (Figure C.14) (Bode, 2003; Goel et al., 2019; Itayama and Sawada, 1995; Pierobon, 2015). We confirmed that 6–12 hr after resection the animals indeed showed loss of specific neuronal cell types by measuring the expression levels of subtype-specific neuronal markers via qPCR (Figure C.15). To limit the stress from microfluidic immobilization that could exacerbate the resection wounds and affect activity, we shortened the duration of these experiments for the majority of the animals (40 min total, 20 min of no stimulation, 20 min of stimulation – valve on 1 s at 22 psi or 0 psi, off 30 s). Only three animals per each condition (stimulated and non-stimulated, and five different resections) were experimented on with a longer duration protocol as used previously (100 min total, 20 min no stimulation, 60 min of stimulation – valve on 1 s at 22 psi or 0 psi, off 30 s, 20 min no stimulation; Figure C.16).

We began resection studies by removing the peduncle and basal disk to create a 'footless' *Hydra.* We hypothesized that aboral neurons may be important for coordinating and enhancing body contractions (Figure C.3). We based this hypothesis on the fact that aboral neuron activity has a strong correlation with body contractions (Badhiwala et al., 2018; Dupre and Yuste, 2017). In addition, the neuropeptide Hym-176C has been shown to induce ectodermal muscle contractions and is selectively expressed in the ectodermal peduncle neurons (Klimovich et al., 2020; Noro et al., 2019; Siebert et al., 2019; Yum et al., 1998). Finally, the presence of gap junction protein innexin-2 in aboral neurons could facilitate fast electrical conductions that allows these neurons to fire synchronously (Siebert et al., 2019; Takaku et al., 2014). This could be necessary for

enhancing neuromuscular signaling for body contractions. Because 'footless' *Hydra* lacked peduncle neurons that we had used previously to measure contraction pulses and bursts (Figure C.16 D-F and Videos C.6–C.9), we performed these experiments using a transgenic *Hydra* line expressing GCaMP7b in the endodermal epitheliomuscular cells, which allowed us to measure contraction pulses and bursts by averaging calcium activity in all the epitheliomuscular cells (using whole-frame ROI, which is more robust to motion artifacts than peduncle ROI; Figures C.3 A-B, C.7, and C.16).

Because neurons in the foot fire synchronously with body contractions, we expected 'footless' animals to show significant changes in contraction behaviors (calcium spiking activity) to be significantly affected by their removal (Badhiwala et al., 2018; Dupre and Yuste, 2017; Shimizu and Fujisawa, 2003), but this was not what we observed. Surprisingly, our experiments with 'footless' animals showed that the aboral nerve ring was not required to regulate spontaneous contraction bursts or pulses or mechanosensory responses. After we removed the peduncle network in *Hydra*, we found that the increase in contraction burst or pulse activity with stimuli (or mechanosensory response) in 'footless' individuals was similar to the increase in activity observed in whole individuals ('footless' N = 8 animals stimulated 20 min, Cohen's d = 1.89, Cliff's delta = 0.81, $p<0.01$; whole N = 8 animals stimulated 20 min, Cohen's d = 1.80, Cliff's delta = 1.00, $p<0.01$; Figure 3b, c). Furthermore, we found no significant difference in either the spontaneous contraction probability or the mechanosensory response probability of 'footless' animals and whole animals ('footless' N = 3 animals not stimulated, spontaneous contraction probability = 0.13 ± 0.01, mean ± SEM; 'footless' N = 3 animals stimulated for 60 min, mechanosensory response probability = 0.88 ± 0.05, mean ± SEM; whole N = 3 animals not

stimulated, spontaneous contraction probability = 0.15 ± 0.01, mean ± SEM; whole N = 3 animals stimulated for 60 min, mechanosensory response probability = 0.73 ± 0.07, mean ± SEM; Figure C.16 B-C and Videos C.4, C.5, C.10, C.11).

Oral neurons play a major role in mechanosensory response in *Hydra*

While the mechanosensory response in *Hydra* remained unaffected with the removal of the aboral nerve ring, we found that removal of the hypostome and tentacles (or 'headless' *Hydra*) resulted in significant changes in both the mechanosensory response and spontaneous contraction bursts or pulses. When we measured the mechanosensory response in 'headless' *Hydra,* we found that the animals still responded to mechanical stimulation with a significant increase in their contraction bursts or pulses; however, they did so with a lower magnitude compared to whole and 'footless' individuals ('headless' N = 8 animals stimulated for 20 min, $p<0.01$, Cohen's d = 1.37, Cliff's delta = 1.00; Figure 3b, c). Specifically, the 'headless' *Hydra* responded with (>2×) lower probability compared to whole animals, and they also showed a (>3×) lower probability of spontaneous contraction bursts and pulses ('headless' N = 3 animals stimulated for 60 min, mechanosensory response probability = 0.29 ± 0.02, mean ± SEM; 'headless' N = 3 animals not stimulated, spontaneous contraction probability = 0.05 ± 0.01, mean ± SEM; Figure C.16 B-C and Videos C.12 and C.13).

To verify the reduction in mechanical response probability was specific to removing the oral network of neurons, and not simply the result of injury, we performed experiments with animals that we cut longitudinally along the body axis ('bisected' *Hydra*) to remove approximately the same amount of tissue while preserving the neuronal subtypes in both the oral and aboral

networks. We found that these longitudinally 'bisected' animals showed mechanosensory responses that were not significantly different from that of the whole, 'headless,' or 'footless' animals ('bisected' N = 3 animals stimulated for 60 min, mechanosensory response probability = 0.54 ± 0.14, mean ± SEM; Figure C.16 B and C). However, the magnitude of increase in contraction bursts or pulses activity due to stimulation in 'bisected' individuals, though lower than whole and 'footless' individuals, was larger than 'headless' individuals ('bisected' N = 8 animals stimulated for 20 min, $p<0.001$, Cohen's d = 2.32, Cliff's delta = 1.0). This suggests that our observations in 'headless' *Hydra* indeed depended upon the types of neurons removed during the headless resection and not simply an injury response. We also found that these longitudinally 'bisected' animals had a lower probability of spontaneous contraction bursts or pulses than the whole animals but higher than 'headless' animals. This suggests that the entire network needs to be intact for normal contraction bursts or pulses activity, and the loss of roughly half the network leads to some reduction in contractile activity ('bisected' N = 3 animals not stimulated, spontaneous contraction probability = 0.08 ± 0.02, mean ± SEM; Figure C.16 B-C and Videos C.14 and C.15).

We next asked if the body column alone is sufficient to mediate the mechanosensory response. To answer this question, we completely removed both oral and aboral regions. In 'body column' animals, we found significant reduction in both the mechanosensory response and spontaneous contraction bursts and pulses relative to whole animals. The 'body column' animals had a mechanosensory response probability that was not different from the mechanosensory response in 'headless' animals, while the spontaneous contraction bursts and pulses probability was lower compared to that of 'headless' animals ('body column' N = 3 animals stimulated for 60 min,

mechanosensory response probability = 0.19 ± 0.08; 'body column' N = 3 animals not stimulated, spontaneous contraction probability = 0.02 ± 0.01 mean ± SEM; Figure 3—figure supplement 3b, c). Although we found a significant increase in contraction bursts and pulses with stimulation as compared to spontaneous contraction bursts and pulses activity in 'body column' individuals similar to all resections, the magnitude of the increase was the lowest in 'body column' animals (even lower than 'headless' animals) ('body column' N = 6 animals stimulated for 20 min, $p<0.05$, Cohen's d = 1.10, Cliff's delta = 0.50; Figure C.3 B-C and Videos C.16 and C.17). Moreover, we did not observe significant increases in contraction bursts and pulses over the same time period (comparing activity from 0 to 20 min with activity from 20 to 40 min, see Materials and methods) in non-stimulated animals, suggesting that the higher probability of contraction bursts and pulses was in fact due to mechanical stimulation (Figure C.17). Thus, based on the comparison between the probability of spontaneous contraction bursts and pulses and mechanosensory response in all resections, we found that the 'body column' animals had a weak response to touch despite their slightly increased contraction bursts and pulses probability with mechanical stimulation.

Oral and aboral networks show different patterns of activity during spontaneous contractions compared to mechanically stimulated contractions

To identify how the activity of neurons in the oral and aboral networks coordinate spontaneous and stimulated responses, we manually tracked the calcium activity of several neurons in the oral and aboral regions (20 min no stimulation, 10 min stimulation 1 s every 31 s, n = 3 *Hydra*; Figures C.19 A-B, C.20, and C.21). We found that there were at least two independent networks of neurons based on a correlation analysis (Figure C.4 C). Specifically, we time-aligned the

calcium activity with either spontaneous contractions or mechanical stimulation events to classify these groups of neurons based on their activity (Figure C.4 D, E, and G). One group of correlated neurons found throughout the entire body showed averaged calcium activity less than 1 s after a mechanical stimulus and spontaneous activity that is consistent with previously reported contraction burst (CB) neurons (Dupre and Yuste, 2017). We plot calcium dynamics of these CB neurons as shades of blue in Figures C.4, C.20, and C21. These neurons show bursts of activity that are synchronized with muscle contractions and show calcium activity that is highly correlated with the average peduncle ROI. In addition to these CB neurons, we found other groups of correlated neurons with average calcium activity that is independent of the CB network. One group showed a distinctive pattern of activity following mechanical stimulation, but no distinctive activity associated with spontaneous contractions. Specifically, this group of neurons near the oral end responded approximately 10 s after mechanical stimulation (Figure C.4 E and G). We found these putative 'mechanically responsive (MR) neurons' in all three *Hydra* we analyzed and plot their calcium dynamics as shades of red in Figures C.4, C.20, and C21. The fact that these MR neurons do not show activity associated with spontaneous contractions clearly indicates that they are not a part of the CB network, but rather these two distinct networks (CB and MR) are involved in the *Hydra's* response to mechanical stimulation. We also note that these MR neurons do not fire periodically as would be expected for the rhythmic potential (RP) network (Dupre and Yuste, 2017). In addition to CB and MR neurons, in *Hydra* 2 (Figure C.19) and 3 (Figure C.20) we also found neurons that were not associated with either the MR network or the CB network. These neurons we labeled as 'unspecified groups' do not appear to be a part of the CB or RP networks previously characterized nor the MR neurons we identify here.

These data suggest that there are at least two separate pathways involved in the mechanosensory response. The first involves the CB neurons and muscle contractions. The second network involving the MR neurons responds more slowly, with a latency of several seconds. Because the activity of the MR neurons occurs after the contraction, their role in the behavioral response remains unclear.

DISCUSSION

Our experiments with 'footless,' 'headless,' 'bisected,' and 'body column' animals show that mechanosensory and spontaneous behaviors are regulated by neural ensembles that are localized to select regions of the animal; however, some properties of the mechanosensory response may be evenly distributed throughout the body. We demonstrate that localized touch produces an increased calcium activity in both peduncle neurons (Figure C.2) and endodermal epitheliomuscular cells (Figure C.3), which is associated with body contractions (Figure C.8) (Badhiwala et al., 2018). We identified at least two neuronal networks (MR and CB network) with distinct neuronal activities mediating the stimulated responses (Figure C.4), where the CB network of neurons show fast calcium responses, and the MR neurons show slower calcium response. It is possible that the MR neurons, primarily found in the oral end (N = 3 *Hydra*), may consist of ec4 neurons, which are the only neuronal subtype in the oral end with an unknown function. The other orally located ectodermal neuron populations, ec1B and ec3C, are suggested to belong to the CB and RP1 circuits, respectively (Siebert et al., 2019). *Hydra's* responsiveness (i.e., probability of response to a mechanical stimuli) depends on the stimulus intensity (Figure C.2). Interestingly, we found that the 'headless' and 'footless' animals can still respond to mechanical stimuli despite missing an entire regional neuronal network; however, 'headless'

animals show reduced responsiveness (Figures C.3 and C.16). The ‘body column’ animals missing both regional neuronal networks have the most dramatically reduced responsiveness, suggesting that these two neuronal networks work together and play compensatory roles in mediating the mechanosensory response.

Surprisingly, although the activity of the peduncle nerve ring is strongly associated with spontaneous contractions, these neurons are not necessary for body contraction and response to mechanical stimulation. This raises the question of what role the peduncle nerve ring plays in *Hydra* behavior. One possible explanation is that this nerve ring coordinates body contractions by enhancing the neural signal to epithelial cells. An additional 4 s reduction in body contraction duration in ‘body column’ animals compared to headless animals supports the idea that the peduncle nerve ring is also involved in coordination of contractile behavior (Figure C.18). Furthermore, calcium activity propagates from the foot to the hypostome in whole animals during body contractions (Szymanski and Yuste, 2019), supporting a hypothesis that the peduncle network of neurons may be motor neurons.

Based on these observations, we propose a simple model for sensorimotor information flow in *Hydra* where we consider the hypostome as an integration point where sensory and motor information converge. The information is then communicated to the peduncle where it is amplified for coordinated whole-body control. There may be sufficient functional redundancy between the hypostome and peduncle regions such that removal or damage to one of them is well tolerated in *Hydra*. Moreover, the diffuse network in the body column may retain minimal processing capabilities needed for weak sensorimotor responses. The fast and slow calcium

responses from the CB and MR neurons, respectively, indeed support the hypothesis that *Hydra* have multiple, separate pathways for behavioral response.

Overall, the quantitative characterization of *Hydra's* sensorimotor responses reported here helps to build the foundation for a more comprehensive investigation of information processing in *Hydra* – an animal with clear advantages to supplement commonly studied model organisms in neuroscience. Important next steps include developing mechanistic models to describe sensory information processing that supports our results. Although our experiments describe general information flow in *Hydra*, having a cellular-level control of neuronal activity would be a clear advantage for revealing the function of neuronal cell types as well as their functional connectivity in the sensorimotor circuits. We expect additional future work with optogenetic manipulation of specific neuronal subtypes combined with fast, volumetric and ratiometric imaging techniques will provide a more comprehensive approach for characterizing this sensorimotor processing in *Hydra*. Recognizing that calcium fluorescence imaging is limited in its ability to measure single spikes (Huang et al., 2021), we also expect other activity sensors (such as voltage indicators) will reveal what role the MR neurons play in behavioral response. Building upon the work reported here, one can then interrogate the roles of these regional networks with multiple sensory modalities, such as light and heat, to answer questions about how diffuse nervous systems may be capable of centralized information processing and multisensory integration. These studies may help reveal a comprehensive model for how internal states and external stimuli shape the behavioral repertoire in an organism with a highly dynamic neural architecture.

MATERIALS AND METHODS

Hydra strains and maintenance

Hydra were raised in *Hydra* medium at 18°C in a light-cycled (12 hr:12 hr; light:dark) incubator and fed with an excess of freshly hatched *Artemia nauplii* (Brine Shrimp Direct, Ogden, UT, #BSEP8Z) three times a week (protocol adapted from Steele lab). All experiments were performed at room temperature with animals starved for 2 days. The transgenic line nGreen, kindly provided by Rob Steele, was generated by microinjecting embryos with a plasmid containing the *Hydra* actin promoter driving GFP expression (Siebert et al., 2019). The transgenic strains expressing GCaMP6s under the actin promoter in neurons and in ectodermal epitheliomuscles (Addgene plasmid: #102558) were developed by microinjections of the embryos by Christophe Dupre in the Yuste lab (Columbia University) (Wicks et al., 1996). The transgenic strain expressing GCaMP7b in endodermal epitheliomuscular cells was co-developed by Juliano Lab (University of California, Davis) and Robinson Lab (Rice University).

Briefly, the plasmid with codon-optimized GCaMP7b under the EF1a promoter was constructed by GenScript (*https://www.genscript.com*). Injections were performed as previously described (Juliano et al., 2014) with the following modifications: (1) injection solution was prepared by mixing 1 µL 0.5% phenol red (Sigma P0290-100ML) with 6 µL plasmid DNA solution prior to centrifugation, and (2) embryos were fertilized for 1–2 hr prior to injection. Plasmid promoters were cloned in expression vector pHyVec2 (Addgene plasmid: #34790) using restriction sites Nsil. Plasmids were prepared by Maxiprep (Qiagen, Valencia, CA) and eluted in RNase-free water. A plasmid DNA solution of 1.4 µg/µL was injected into embryos using an Eppendorf FemtoJet 4x and Eppendorf InjectMan NI 2 microinjector (Eppendorf; Hamburg, Germany)

under a Leica M165 C scope (Leica Microscopes, Inc; Buffalo Grove, IL). Viable hatchlings with mosaic expression were propagated by asexual reproduction, and asexual buds were screened and selected for increasing amounts of transgenic tissue until a line was established with uniform expression in the endodermal epithelial cells.

Fluorescence imaging of *Hydra* nerve net

Distribution of neurons in the *Hydra* nerve net was fluorescently imaged with transgenic *Hydra vulgaris* expressing GFP (nGreen) in neurons and neuronal progenitors (Figure C.1 A; (Siebert et al., 2019). *Hydra* was anesthetized with 0.05% chloretone and immobilized in an ~160 μm tall microfluidic chamber (Badhiwala et al., 2018). High-resolution fluorescence imaging was performed using a confocal microscope (Nikon TI Eclipse) and 10× (0.45 NA) objective, where the *Hydra* was imaged at a single plane with multiple fields of views stitched together to obtain an image of the whole animal (Figure C.1 A).

***Hydra* resections**

Hydra were placed in a Petri dish and covered with enough *Hydra* medium to prevent desiccation. When *Hydra* were relaxed and stationary, resections were performed with a single incision with a scalpel across the body. We referenced published images of in situ hybridizations of different cell types and spatial expression patterns to guide the location of incisions. For 'footless' *Hydra,* an axial cut above the peduncle removed approximately one third of the lower body, which included the peduncle and the basal disk. For 'headless' *Hydra,* an axial cut below the hypostome removed approximately one third of the upper body, including the tentacles and the hypostome. For 'bisected' *Hydra*, a transverse cut starting from the tip of the hypostome to

the basal disk was made along the midline of the body. For 'body column' *Hydra*, an axial cut above the peduncle removed the lower body followed by another axial cut below the hypostome to remove the upper body region. This preparation resulted in an open tube body. *Hydra* were stored in an 18°C incubator after the excisions and until beginning the experiments. *Hydra* can seal open wounds within ~3–4 hr and repopulate the neuronal population to regain functionality in 30–72 hr (Figure C.14, see 'Imaging regeneration of peduncle network'). To allow *Hydra* time to recover but not regain the functionality of lost neuronal cell types, we performed experiments 6–12 hr post amputation.

Imaging regeneration of peduncle network

Transgenic *Hydra* (GcaMP6s, neurons) were axially cut in the middle of the body column to generate an oral and aboral halves (Figure 3.14 A). The oral half was immobilized between two coverslips with an ~110 μm PDMS spacer. Calcium fluorescence was conducted for 20 min every 2 hr on Nikon TI Eclipse inverted microscope with 20% excitation light from Sola engine and GFP filter cube. We captured frames at ~10 Hz (100 ms exposures) with Andor Zyla 4.2 sCMOS camera with NIS software. We used 4× (0.2 NA) objective for wide-field imaging to fit the entire *Hydra* in the field of view reduce the likelihood of *Hydra* migrating out of the imaging frame. *Hydra* were exposed to blue excitation light for 20 min during imaging and remained under dark conditions for 100 min between subsequent imaging timepoints. One animal was imaged for ~20 hr starting with 1 hr post resection (Figure C.14 B). *Hydra* had a visibly open wound at the first imaging point but was not detectable after 3 hr. There were no visibly active neurons in the regenerating aboral end, indicating that the resected peduncle neurons had not regenerated. Another animal was imaged for ~30 hr starting with 37 hr post resection (Figure

C.14 C). At the first imaging timepoint (t = 37 hr post resection), there were few neurons in the peduncle region that were active during body contractions, and these groups of neurons resembled a nerve ring as early as 41 hr post resection. Although the peduncle nerve ring seemed to have formed at this point, it appeared to not be as densely populated qualitatively as observed in whole animals.

Microfluidic device fabrication

The mechanical stimulation devices are double-layer microfluidic devices with push-down valves custom-designed with CAD software (L-edit) and fabricated using standard photo- and soft-lithography techniques. All master molds were fabricated with transparency photomasks and SU-8 2075 (MicroChem). The master mold for the bottom *Hydra* layer (circular chambers, 3 mm diameter) was fabricated with the height of ~105–110 μm thick pattern (photoresist spun at 300 rpm for 20 s, 2100 rpm for 30 s). The master mold for the top valve layer (nine individual circular valves, 3 × 3 arrangement, 400 μm diameter each) was fabricated with height of ~110 μm thick (photoresist spun at 300 rpm for 20 s, 2100 rpm for 30 s). Polydimethylsiloxane (PDMS) Sylgard 184 was used to cast microfluidic devices from the master molds. The bottom *Hydra* layer (10:1 PDMS spun at 300 rpm for 40 s ~3 hr post mixing; cured for 12 min at 60°C) was bonded to the valve layer (~4 mm thick, 10:1 PDMS; cured for ~40 min at 60°C with holes punched for inlet ports) with oxygen plasma treatment (Harrick Plasma, 330 mTorr for 30 s) and baked for at least 10 min at 60°C. *Hydra* insertion ports were hole-punched through the two layers for *Hydra* layer with 1.5 mm biopsy punches, and the devices were permanently bonded (O_2 plasma treatment, 330 mTorr for 30 s) to 500 μm thick fused silica wafer (University Wafers) and baked for at least 1 hr at 60°C. The design files for the photomasks and step-by-step

fabrication protocols will be available on *https://www.openHydra.org* (under Resource hub/Microfluidics).

Hydra were immobilized and removed from the microfluidic device using syringes to apply alternating positive and negative pressures as previously reported (Badhiwala et al., 2018). The microfluidic devices were reused after cleaning similarly to the protocol previously reported. Briefly, the devices were flushed with deionized water, sonicated (at least 10 min), boiled in deionized water (160°C for 1 hr), and oven-dried overnight.

After repeated use, the PDMS stiffness can change and affect the valve deflection and the actual force experienced by *Hydra* through the PDMS membrane. Additionally, uncontrollable conditions during the device fabrication process can also lead to small differences between devices. As a result, all data for Figure 2 were taken with a single device and the response curve was used to calibrate (identify pressure that yielded ~60% response probability equivalent to 20–22 psi stimulus intensity) new devices.

Microfluidic mechanical stimulation

We used compressed air to inflate the microfluidic valves. For temporal control over valve (on/off) dynamics, we used a USB-based controller for 24 solenoid pneumatic valves (Rafael Gómez-Sjöberg, Microfluidics Lab, Lawrence Berkeley National Laboratory, Berkeley, CA 94720) and a custom-built MATLAB GUI (available at *https://www.openHydra.org*) that allowed setting the stimulation parameters, such as the duration of valve ‘on’ (1 s), duration of valve ‘off’ (30 s), and the duration of stimulation (60 min, 119 cycles of valve ‘on’ and ‘off’),

and pre, post-stimulation acclimation/control period (20 min). We used a pressure regulator to manually control the air pressure into the valve manifold. In summary, we set the stimulation pressure with a pressure gauge to regulate the flow of air into the valve manifold. This valve manifold was controlled with a USB valve controller that allowed us to programmatically inflate the valve (turn it 'on') with pressurized air with custom stimulation parameters.

To test sensory motor response to mechanical stimuli, we used pressurized air to inflate the push-down valve and cause it to press down on the *Hydra* immobilized in the bottom layer. Each experimental condition had at least three *Hydra* each. For a given condition, replication experiments were conducted on different days. After an animal was immobilized inside the *Hydra* chamber, we selected one valve (from the nine valves over the entire chamber) that was directly above the midbody column region to deliver stimuli. Although *Hydra* were free to move, we did not observe large displacement most of the time, and, as a result, the same valve remained in contact with the animal throughout the stimulation period.

We adjusted the air pressure using a pressure regulator for each experiment, and the valves were inflated using a USB valve controller (see above). The full-length stimulation experiment consisted of 20 min of no stimulation (control/acclimation) followed by 60 min of stimulation period (except habituation experiment where the stimulation period was 120 min), where valves were pulsed with constant pressure (0 [control], 5, 10, 15, 20, 22, or 25 psi) for 1 s every 31 s, then another 20 min of no stimulation (control/acclimation). Shorter stimulation experiment (used for whole-animal muscle imaging of various resections) consisted of 20 min of no stimulation (acclimation/control) followed by 20 min of stimulation period (valves pulsed for 1 s

every 31 s with a constant pressure of 0 or 22 psi). We chose a 20 min initial control period based on the high sensitivity to abrupt changes in light intensities (especially to blue wavelengths used for excitation of GCaMP) in *Hydra,* which leads to increased contractile activity for 2–5 min. Even with the stimulus repeated for 2 hr at a constant inter-stimulus interval, we found no obvious evidence of sensitization, habituation, or stimulus entrainment in *Hydra* (Figure C.13). As a result, we chose not to randomize the inter-stimulus interval.

Distribution of mechanical forces

We characterized the distribution of force exerted by a microfluidic valve by quantifying the movements of neurons due to mechanical stimulation. We performed fluorescence imaging in transgenic *Hydra* (nGreen) expressing GFP in neurons for ~8 min. We captured 8000 frames at ~16 Hz (50 ms exposures) with 4× objective (0.2 NA) and Andor Zyla 4.2 sCMOS camera with 2 × 2 binning (1024 × 1024 frame size) using MicroManager. *Hydra* was stimulated in the middle of the body column (valve pulsed for 1 s every 31 s, five times).

To quantify the movement of neurons and tissue throughout the *Hydra* body, we tracked a total of 222 neurons that were visible for 1 min capturing the first two stimulation trials. We performed semi-automated tracking with TrackMate plugin (ImageJ/Fiji) (Rueden et al., 2017; Schindelin et al., 2012; Tinevez et al., 2017) and manually corrected the tracks where neurons were misidentified. From these tracks, we calculated the displacement of each of the neurons between each frame (~50 ms). The average cellular displacement (0.4 μm per frame, 50 ms) was calculated by averaging the cellular displacements from all frames when the valve was not pressurized. We found a significantly increased displacement (6.2 μm, $p<0.001$) just after the

valve was pressurized (and after the valve was depressurized 1 s later). We then generated a vector map of neuronal displacements by calculating the change in position of each of the neurons in the frame just after the valve was pressurized for the first stimulation trial (Figure C.5 A). We also plotted the cellular displacement for each of the neurons and the location of those neurons relative to the center of the valve. We averaged the highest three displacements in 50 μm radial band increments from the valve center to quantify how far the mechanical forces extended. This was a more conservative measurement as the neurons in different tissue layers (endodermal and ectodermal layers furthest from the valve) may have experienced different forces. By taking the average of the highest three displacements in 50 μm radial bands increments from the valve center, we identified the majority of the (shear) force was experienced by neurons bordering the valve in 250 μm radius.

Fluorescence intensity from GFP *Hydra*

To prove that the fluorescence intensity changes we observed with calcium imaging are due to calcium activity and not motion artifacts from body contractions, we compared the average fluorescence intensities from three different transgenic *Hydra* lines: (1) expressing GFP pan-neuronally (nGreen line), (2) expressing GCaMP6s pan-neuronally, and (3) expressing GCaMP7b in ectodermal epitheliomuscular cells (Figure 2—figure supplement 3). For each *Hydra*, we compared the average fluorescence intensities from three different ROIs that included the peduncle region, whole-frame (for whole body) region, and valve region. We also measured the body length by taking the major axis of an ellipse fitted along the oral-aboral axis of the body column (from apex of the hypostome to the basal disc) after binarizing the fluorescence image.

For this analysis, we used fluorescence imaging from transgenic *Hydra* expressing GFP pan-neuronally (nGreen) performed in 'Distribution of mechanical forces.' *Hydra* was stimulated in the middle of the body column (valve pulsed for 1 s every 31 s, five times). We used calcium imaging from transgenic *Hydra* (expressing either GCaMP6s pan-neuronally or GCaMP7b in endodermal epitheliomuscular cells) stimulated in the middle of the body column (valve pulsed for 1 s every 31 s, 120 times). The average fluorescence intensity and body length traces were time-aligned to the onset of mechanical stimulation to show (1) changes in average fluorescence intensity in peduncle ROI and the whole-frame ROI are due to calcium activity (increase in fluorescence from in GCaMP lines) and not motion artifact (no change in fluorescence from GFP line) and (2) increase or decrease in body length or average fluorescence from valve ROI are affected by stimulation artifacts.

Average calcium fluorescence from large and small ROIs in the peduncle

We performed fluorescence imaging in transgenic *Hydra* expressing GCaMP6s in neurons for ~1 min during spontaneous behaviors (*Hydra* was not stimulated). We captured frames at ~16 Hz (50 ms exposures) with a 10× objective (0.45 NA) and Andor Zyla 4.2 sCMOS camera with 3 × 3 binning (682 × 682 frame size) using MicroManager.

Because calcium fluorescence decreases when neurons are inactive, tracking multiple neurons in a highly deformable region is a challenge. Nonetheless, we tracked nine neurons from the peduncle nerve ring that were visible (enough to track) for 1 min. We performed semi-automated tracking with TrackMate plugin (ImageJ/Fiji) (Rueden et al., 2017; Schindelin et al., 2012; Tinevez et al., 2017) and manually corrected the tracks where neurons were misidentified. We

also used a large peduncle ROI, similar to how we used an ROI for quantifying neural response to stimulation in all other experiments. We calculated the average calcium fluorescence trace from small ROIs for individual neurons and large ROI for the peduncle region. Fluorescence intensity was normalized by calculating $\Delta F/F$, where $\Delta F = F\ F_0$ and F_0 is the minimum fluorescence from prior timepoints. The large peduncle ROI had highly correlated calcium fluorescence activity with smaller ROIs for individual neurons, which are known to belong to the contraction burst circuit (Figure C.6). As a result, for all other experiments we use large peduncle ROI when measuring the neuronal activity.

Whole-animal imaging of neural activity at different stimulus intensities

We characterized the neural response to repeated local mechanical stimuli at 0 (control), 5, 10, 15, 20, and 25 psi pressure with animals that expressed GCaMP6s in neurons (Figure 2). Each pressure condition was experimented with three animals on different days using the same device and stimulation protocol (valve pulsed for 1 s every 31 s for 60 min) to generate the pressure-response curves (total of 18 animals). Calcium fluorescence imaging for all experiments was conducted for the 100 min duration of the stimulation protocol (see Mechanical stimulation subsection) on Nikon TI Eclipse inverted microscope with 20% excitation light from Sola engine and GFP filter cube. We captured 100,000 frames at ~16 Hz (50 ms exposures) with Andor Zyla 4.2 sCMOS camera with 4 × 4 binning using MicroManager. For imaging neural activity, 12-bit low-noise camera dynamic range was used. To synchronize the stimulation onset times with imaging, we use used a data acquisition device (LabJack U3) to record the TTL frame out signal (fire-any, pin #2) from the Zyla and the valve on/off timestamps from the valve controller.

We used 4× (0.2 NA) objective for wide-field imaging to fit the entire *Hydra* in the field of view to reduce the likelihood of *Hydra* migrating out of the imaging frame. There are neurons in the oral region (hypostome) that are co-active with aboral region (peduncle) neurons during body contractions; however, neurons in the hypostome appeared to be much sparser than those in peduncle (Video 18). Discerning these hypostomal network neurons required higher magnification, which significantly reduced the field of view such that considerable amount of nervous tissue could move in and out of the imaging plane (z-plane) or frame (xy-plane), making it difficult to obtain reliable calcium fluorescence time series.

Whole-animal imaging of neural activity with different stimulation regions

To map the sensitivity of different body parts to mechanical stimuli, we performed calcium imaging of multiple animals (N = 3, whole animals expressing GCaMP6s in neurons) using imaging settings previously described with modified experimental protocol. Based on the range animal sizes (1–2.5 mm) and the size of the valve (400 μm), we stimulated *Hydra* body at three different regions along the oral-aboral axis: (1) the oral end (upper third of the body), (2) the aboral end (lower third of the body), and (3) the third near the midbody column. Each animal was stimulated at three different locations (20 min no stimulation, stimulation region #1 – 1 s every 31 s for 20 min, ~2 min no stimulation, stimulation region #2 – 1 s every 31 s for 20 min, ~2 min no stimulation, stimulation region #3 – 1 s every 31 s for 20 min using 22 psi). We randomized the order in which the three different body regions were stimulated in each of the three animals to avoid any stimulus entrainment artifacts. We then analyzed the peduncle nerve ring activity in response to mechanical stimulation of different body regions as detailed in 'Analysis of calcium activity' and 'Analysis of mechanosensory response'.

Whole-animal imaging of neural activity in resected animals

For experiments that involved body lesions (Figure C.16), the same experiment protocol and imaging settings as previously described (see 'Whole-animal imaging of neural activity at different stimulus intensities,' 20 min no stimulation, stimulation 1 s every 31 s for 60 min, 20 min no stimulation) were used with either 0 psi (control) or 20–22 psi (~60% response probability). Each condition group had three animals (total of 18 animals). The 'headless' and 'bisected' animals were prepared with the appropriate body regions removed as described previously (*'Hydra* resections'). Due to difficulty in tracking neurons and weak GCaMP expression in body column neurons, we were unable to quantify responses from 'footless' and 'body column' *Hydra* and thus excluded them from experiments.

Simultaneous electrophysiology and calcium imaging of ectodermal epitheliomuscular cells

Electrical activity from the epitheliomuscular cells was measured simultaneously with calcium imaging of the ectodermal epitheliomuscular cells in transgenic *Hydra* using a nano-SPEARs device previously reported (Figure C.8) (Badhiwala et al., 2018). Briefly, transgenic *Hydra* expressing GCaMP6s in the ectodermal epitheliomuscular cells starved for at least 48 hr were used to measure the activity of the muscles (10 fps, 30 min, 4× objective with 15% light intensity). An inverted microscope with GFP filter and Andor Zyla 4.2 were used for capturing images. All electrical data was obtained with an Intan Technologies RHD2132 unipolar input amplifier (*http://intantech.com*) at a sampling rate of 1 kHz, low-frequency cutoff and DSP filter of 0.1 Hz and high-frequency cutoff of 7.5 kHz. From the calcium activity traces, we identified 30 s intervals of either high-amplitude activity or low-amplitude activity to perform cross-correlation analysis. The high- and low-amplitude activity regions were manually identified with

a threshold of 20% of the highest peak in the calcium activity. The high-amplitude activity region occurred during contraction bursts. The low-amplitude activity region occurred during tentacle contractions for muscular activity imaging. Both the Intan amplifier and the Zyla were triggered with the same TTL signal. However, to account for any offset in the timing of the electrical and optical data, we measured the maximum of the cross-correlogram in a 50 ms (approximately one duty cycle of the trigger signal) window rather than the cross-correlation at zero offset to generate the correlation maps. For correlation map, each frame was down sampled to 64 × 64 pixels, and the fluorescence trace for the downsampled pixels across the 30 s intervals was cross-correlated with electrical activity during the same 30 s interval. The intensity of color in the correlation map was used to indicate correlation values.

Whole-animal imaging of epitheliomuscular activity at different stimulus intensities

We characterized the muscle response to mechanical stimuli using animals that expressed GCaMP7b under the EF1a promoter in endodermal epitheliomuscular cells (Figure C.3). We measured contraction pulses and contraction bursts by averaging calcium activity in all the epitheliomuscular cells (Figure C.3 A). The correlation between peduncle nerve ring activity and muscle contractions has been previously established based on simultaneous electrophysiology and neuronal (Badhiwala et al., 2018) or ectodermal epitheliomuscular cells calcium imaging (Figure C.8). The imaging protocol for epitheliomuscular cells was similar to the imaging of neural activity (see ‘Whole-animal imaging of neural activity at different stimulus intensities’), except a 16-bit camera dynamic range was used to avoid saturating the sensor.

We first developed a partial psychometric curve for calcium activity of epitheliomuscular cells to confirm the dependence of response on stimulus intensity. Three animals were imaged for 20 min without stimulation and then stimulated at three different pressures (20 min at 15 psi, 20 min at 20 psi, 20 min at 25 psi; or in reverse order), and the epitheliomuscular response curve was used to identify the stimulus intensity (~22 psi), which yielded ~60% response probability (Figure C.9). This stimulation intensity was also selected for stimulation of N = 8 whole animals (40 min of imaging; 20 min of no stimulation, 20 of stim for 1 s every 31 s with 22 psi) and animals with different body regions removed.

Whole-animal imaging of epitheliomuscular activity in resected animals

The various body regions were removed as described previously (*'Hydra* resections') to prepare 'headless,' 'footless,' 'bisected,' and 'body column' *Hydra*. Note that this transgenic line of *Hydra* expressing GCaMP7b in the endodermal cells showed some signs of deficit. They were particularly sensitive to being handled and were more likely to dissociate after ~30 min in the chambers during long-term microfluidic immobilization and fluorescence imaging. The resected *Hydra* were especially difficult to image for the entire 100 min without any cell dissociation. This could be due to the specific promoter used for driving GCaMP expression or that phototoxicity is higher when there is a high expression of GCaMP in a larger number of cells. Because the lesioned animals were more likely to be damaged during microfluidic immobilization, the mechanical stimulation protocol (see Mechanical stimulation subsection) was shortened to total of 40 min of imaging with 20 min of no stimulation followed by 20 min of repeated stimulation (1 s every 31 s for 20 min). For three animals per each resection (total of 30 animals), we used the full-length protocol (100 min of imaging: 20 min of no stimulation, 60 min

of stimulation every 31 s, and 20 min of no stimulation) to obtain higher statistical power for quantifying response probability (Figure C.16). A total of 6–8 animals were stimulated for each resection condition (whole, N = 6 not stimulated, N = 8 stimulated; 'footless,' N = 4, not stimulated, N = 8 stimulated; 'headless,' N = 5 not stimulated, N = 8 stimulated; 'bisected,' N = 5 not stimulated, N = 8 stimulated; 'body column' N = 3 stimulated, N = 6 stimulated; total 61 animals, Figure C.3; 30 of which were experimented with long [100 min] stimulation protocol; Figure C.16).

Analysis of calcium activity

To analyze neural response, we used the average fluorescence from the peduncle region. Tracking individual neural responses was difficult due to high deformability of the body and lack of fluorescence markers when neurons were not active; as a result, we looked at the synchronous firing activity of the neurons in the peduncle region, which is known to have high correlation with body contractions. To analyze epitheliomuscular response, we used average fluorescence from the whole animal to obtain the fluorescence signal over time and analyzed these signals similarly to neural response. Briefly, using ImageJ (Fiji) (Rueden et al., 2017; Schindelin et al., 2012) we used an ROI over the peduncle region or the whole *Hydra* to obtain fluorescence signal over time for neuronal or epitheliomuscular calcium activity, respectively.

From the fluorescence signal, we detected the large calcium spikes as individual contractions (contraction pulses or bursts) using MATLAB peak finding algorithm. We generated a raster plot of calcium spiking activity time-aligned with the stimulus where each row represented one

stimulation trial with 15 s before and after stimulation onset. These raster plots were used to calculate the probability for spontaneous contraction and mechanosensory response.

For contractile behavior analysis, we then annotated the calcium spikes. Single-calcium spikes were labeled as single-pulse contraction. A volley of calcium spikes was labeled as a contraction burst. Both of these led to behavioral contractions, thus time between contractions was calculated as the time between the end of a contraction event (single pulse or burst) and the start of the next one (as shown in Figure C.18). Percent of contractions that are single-pulse contractions was calculated as the fraction behavioral contractions that were single-calcium spikes (not bursts). Contraction duration was used to indicate the amount of time contraction behavior lasted (time from rise in fluorescence signal to return to baseline, as shown in Figure C.18).

Analysis of mechanosensory response

To obtain the response probability (whether an animal had a contraction pulse – either a single pulse or a pulse from a burst) at the time of stimulation (when valve was pressurized), we generated a raster plot of fluorescence spikes time-aligned to stimulus onset and superimposed for each 30 s intervals (time between stimulus). We used an hour of activity (t = 20–80 min) to generate the raster plot and calculate response probability. We defined the 1 s window while the valve was pressed as the response window for each trial (Figure C.10). We then calculated the fraction of all trials (119 trials over 60 min or 40 trials over 20 min) per animal that had at least one fluorescence spike (‘contraction pulse’) in the 1 s response window following stimulus onset to obtain the contraction probability. Extraction of raw fluorescence for ROIs was performed with ImageJ (Fiji) (Rueden et al., 2017; Schindelin et al., 2012) and postprocessing analysis was

performed with MATLAB (using peak-finding algorithm to detect spikes). A one-way ANOVA with Bonferroni correction was used for statistical analysis.

For animals with shorter experiment duration, the mechanosensory response probability in stimulated animals was calculated over the 20 min segment (time = 20–40 min) with a fraction of all trials (40 trials over 20 min) that had at least one fluorescence spike. A one-way ANOVA with Bonferroni correction was used for statistical analysis when comparing multiple conditions. A paired t-test was used to compare the difference between probability of spontaneous contraction and mechanosensory probability in the same animals for each of the conditions. The effect size (magnitude of difference) was measured with Cohen's d and Cliff's delta.

Analysis of spontaneous contraction

For non-stimulated animals, fluorescence activity from 1 hr of activity (time = 20–80 min) was converted into a raster plot with multiple (119 trials over 60 min) 30 s long intervals (to match the stimulation interval in stimulated animals). We obtained 'stimulation times' using a DAQ to record the signal from valve controller except the air pressure was set to 0 psi. The spontaneous contraction probability was calculated by taking the average of response probability from a random 1 s interval (Figures C.3 and C.16). Briefly, by sliding a 1 s window by ~0.3 s over the x-axis (30 s of stimulation interval) on the raster plot (pooled from all three animals for Figure 4a, per individual stimulated animal for Figure C.16), we generated the distribution of the fraction of trials (119 trials × 3 animals over 60 min) with at least one contraction event in a random 1 s response window (Figure C.16). The distributions were compared with a Kruskal–Wallis test.

For animals with shorter experiment duration, spontaneous contraction probability in stimulated animals was calculated similar to non-stimulated animals above, except the raster plot was generated with the first 20 min of fluorescence activity (time = 0–20 min) when no stimulation was applied. Briefly, we generated a distribution of random probabilities by sliding a 1 s window over the x-axis on the raster plot and used the distribution mean as the spontaneous contraction probability to compare with mechanosensory response probability from the same animal with a paired t-test (Figure C.3). For non-stimulated animals with shorter experiment duration, spontaneous contraction probability was also calculated over the second 20 min interval (t = 20–40 min) to confirm the increase in contraction probability during mechanical stimulation was in fact due to stimuli and not just from temporal variation in spontaneous contractions (Figure C.17). Note that these experiments were performed with *Hydra* constrained to ~110 μm thick microfluidic chambers. Although animals are able to behave under such confinements, the range of behavioral motifs and rates may be altered due to compression.

Neuron subtype gene expression analysis with RT-qPCR

Resected *Hydra* were prepared as described above, with 12 polyps per biological replicate (except six whole polyps for control) and a total of two biological replicates per treatment. Approximately 12 hr post resections, tissue was frozen in 1 mL Trizol at −80°C until use. RNA was isolated using the Zymogen RNA Clean and Concentrator kit (Zymogen #R1017) with an in-column Zymogen DNAse I digestion (Zymogen #E1010) following the manufacturer's protocol. cDNA was synthesized using 1 μg of purified RNA and Promega M-MLV RNase H Minus Point Mutant Reverse Transcriptase (Promega, Madison, WI M3682) using the

manufacturer's protocol for oligo dT-primed synthesis. cDNA synthesis was validated via PCR, and all cDNA samples were diluted 1:10 in nuclease-free water for use in qPCR experiments.

Each sample was run in three technical replicates per gene in a 10 µL qPCR reaction using Bio-Rad SsoAdvanced universal SYBR green master mix (Bio-Rad, Hercules, CA, 1725271). Samples were run on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad 1855195). Data were analyzed with the $2^{\wedge-\Delta\Delta Ct}$ method using the 'tidyverse' package in R (Livak and Schmittgen, 2001; Wickham et al., 2019).

Briefly, Cq values from technical replicates were pooled for subsequent analyses. *rp49* was used as an internal control to calculate ΔCq values after first being found to give similar results across all treatments when compared to a second control gene, *actin.* All results were normalized to the 'whole' animal samples. A template-free water control was performed for all primer sets to ensure contamination-free reactions. All qPCR primers (Table C.1) were validated via a 10-fold serial dilution standard curve to have a binding efficiency over 90%.

Whole-animal imaging with single-neuron resolution

We performed fluorescence imaging in transgenic *Hydra* (N = 3) expressing GCaMP6s in neurons for ~30 min (30 min of imaging: 20 min of no stimulation, 10 min of stimulation every 31 s). We captured frames at ~16 Hz (50 ms exposures) with a 10× objective (0.45 NA) and Andor Zyla 4.2 sCMOS camera with 3 × 3 binning (682 × 682 frame size) using MicroManager.

Because calcium fluorescence intensity is low (almost indiscernible from the autofluorescence of *Hydra*) when neurons are inactive, tracking multiple neurons in a highly deformable region without static nuclear fluorescence (such as RFP) is a challenge. Nonetheless, we manually tracked 5–6 neurons throughout the animal body that were visible (enough baseline fluorescence to track even when the neurons were inactive) for the entire duration of the imaging. We performed semi-automated tracking with TrackMate plugin (ImageJ/Fiji) (Rueden et al., 2017; Schindelin et al., 2012; Tinevez et al., 2017) (to track individual neurons with small circular ROIs (~5 μm radius). We manually tracked (interpolating the current ROI location based on the past and future locations) the neurons when the fluorescence levels were dimmer than the background autofluorescence or when neurons were misidentified by TrackMate. We also used a large peduncle ROI, similar to how we used an ROI for quantifying neural response to stimulation in all other experiments. We calculated the average calcium fluorescence traces from small ROIs for individual neurons and large ROI for the peduncle region. Each of the fluorescence traces were corrected by calculating $\Delta F/F_0$, where $F_0$is the mean fluorescence intensity of the trace.

We performed the correlation analysis (MATLAB) of the calcium fluorescence time series for the individual neuron ROIs and peduncle ROI to identify groups of neurons with correlated activity. We calculated the correlation coefficients for the individual neurons and the peduncle ROI by correlating the entire 30 min calcium fluorescence trace for each of the neurons. These correlation coefficients were shown to be statistically significant then randomly reshuffled calcium fluorescence. We divided each of the neuronal calcium time series into 30 blocks. These individual blocks were randomly recombined to generate reshuffled time series for each of the

neurons to calculate correlation coefficients. This random reshuffling of each neuron was repeated 1000 times to obtain the mean correlation coefficients for each of the neuron pairs (Figure C.21 D-F). We then calculated the z-score $(x-\mu)/\sigma$, where x is the correlation coefficients of the original raw time series, μ is the mean correlation coefficients of the time series randomly reshuffled 1000 times, and σ is the standard deviation of the correlation coefficients of the time series randomly reshuffled (Figure C.21 G-I). To identify the neuronal clusters, we used a hierarchical clustering algorithm (MATLAB function 'linkage' with Euclidean distance of correlation coefficients) on the correlation coefficients. We then used leaf order from the linkage tree to sort the neurons and generate the correlation heat map.

We calculated the average fluorescence traces from the time-aligned calcium fluorescence to either spontaneous contractions or stimulus events to identify the roles of these neurons. For spontaneous contractions, the calcium fluorescence (t = 0–20 min) was time-aligned to the onset of contraction bursts or pulses from the peduncle ROI. For stimulated contractions, the calcium fluorescence (t = 20–30 min) was time-aligned to the onset of stimulus recorded with a DAQ (see 'Whole-animal imaging of neural activity at different stimulus intensities'). We further confirmed the different neuronal clusters identified by performing correlation analysis on all neurons from all three *Hydra* by calculating the correlation coefficients for the average fluorescence traces from the time-aligned calcium fluorescence to stimulus events (Figure C.4 F-G).

FIGURES

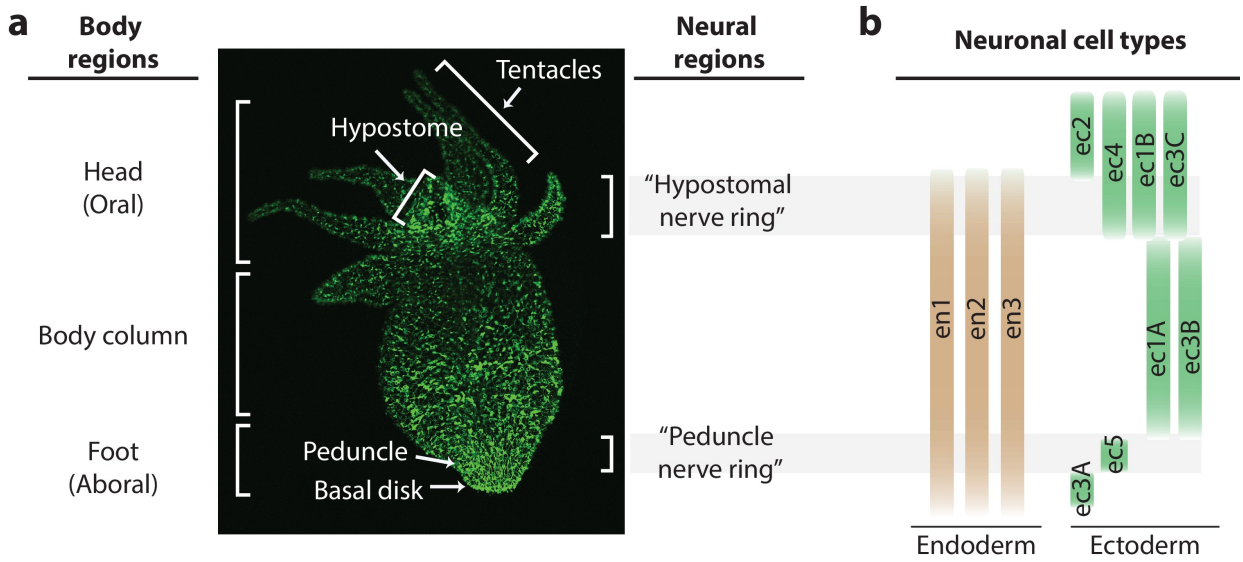
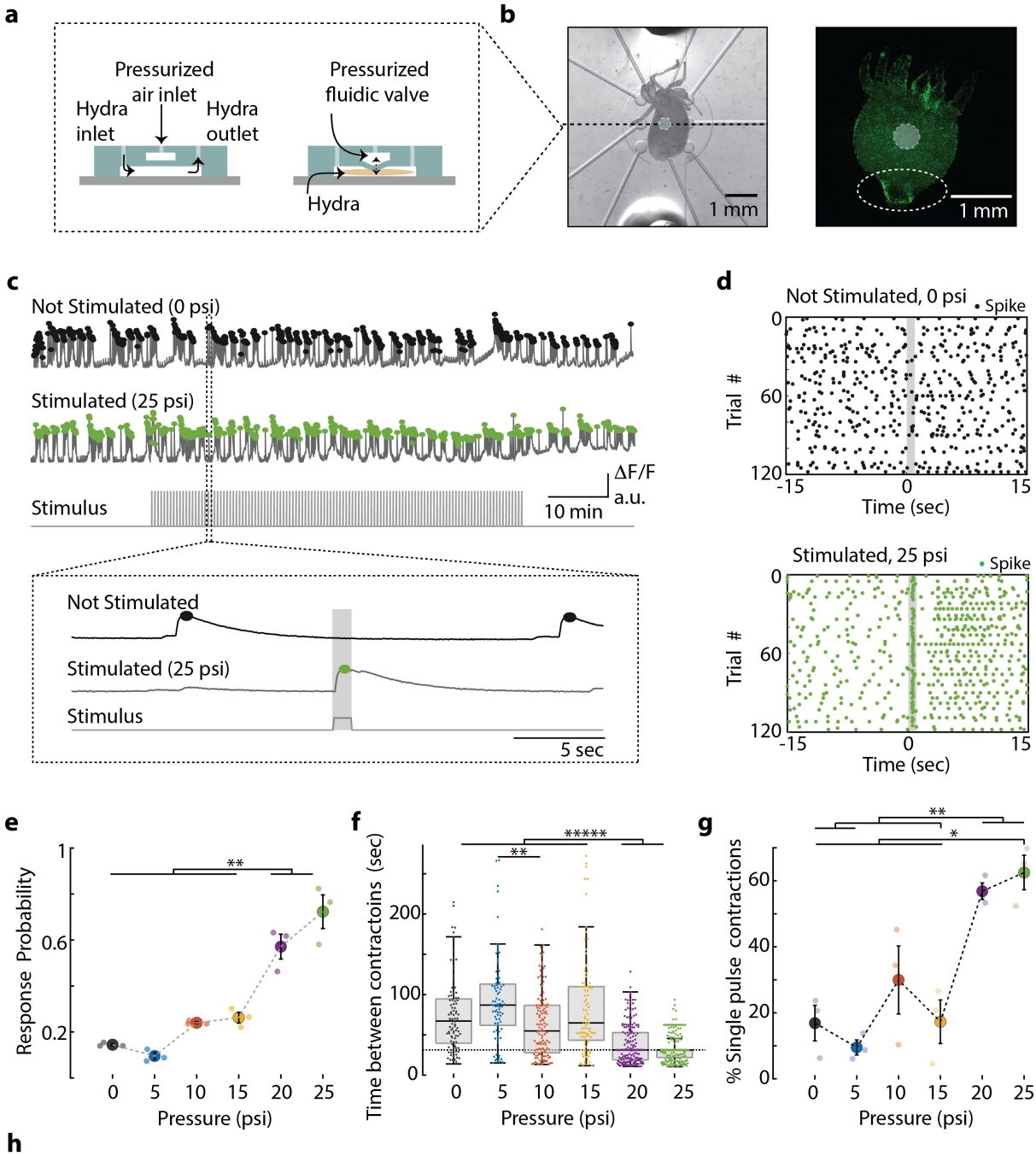


Figure C.1. Distribution of neurons in the *Hydra* nerve net. (A) Fluorescent image of *Hydra* nervous system. Green fluorescent protein (GFP) is expressed in neurons and neuronal progenitors (nGreen transgenic line; (Siebert et al., 2019). Body anatomy is annotated on the left. White arrows indicate the body parts: hypostome, tentacles, peduncle, and basal disk. High neuronal density regions are annotated on the right. (B) Distribution of neuronal cell types varying longitudinally along the body, with endodermal nerve net cell types in tan and ectodermal nerve net cell types in green. Cell types were identified through single-cell RNA sequencing (Siebert et al., 2019).



h

Pressure (psi)	Response probability mean +/- SEM	Time between contractions mean +/- SEM	median +/- SE	Single pulse contractions mean +/- SEM
0	0.11 +/- 0.01	72.98 +/- 4.58 sec	66.68 +/- 5.73 sec	16.87 +/- 5.38 %
5	0.09 +/- 0.02	92.66 +/- 5.87 sec	86.65 +/- 7.35 sec	9.52 +/- 2.27 %
10	0.23 +/- 0.01	63.01 +/- 3.73 sec	54.73 +/- 4.67 sec	29.96 +/- 10.29 %
15	0.24 +/- 0.03	83.60 +/- 6.37 sec	64.73 +/- 7.98 sec	17.30 +/- 6.59 %
20	0.60 +/- 0.06	38.28 +/- 1.77 sec	31.14 +/- 2.21 sec	56.84 +/- 2.48 %
25	0.77 +/- 0.08	32.20 +/- 1.04 sec	31.14 +/- 1.30 sec	62.50 +/- 5.23 %

Figure C.2. *Hydra*'s neuronal response depends on the mechanical stimulus intensity. (A) (Left) Side view of double-layer microfluidic device for mechanical stimulation. (Right) Device with pressurized valve. *Hydra* is immobilized in the bottom *Hydra* layer, and pressurized air supplied into the valve layer causes the circular membrane (400 μm diameter) to push down on *Hydra*. (B) (Left) Brightfield image of *Hydra* immobilized in the bottom layer of the chip and the arrangement of micro-valves on the top layer. The micro-valve used for stimulation is falsely colored with a light blue circle. (Right) Fluorescent image of *Hydra* with pan-neuronal expression of GCaMP6s. White dashed circle marks the peduncle region of interest (ROI) used for quantifying calcium fluorescence changes. (C, D) Representative calcium fluorescence activity in the peduncle region from an animal not stimulated and an animal mechanically stimulated with 25 psi. Black and green dots indicate fluorescence (calcium) spikes. Gray shaded regions indicate stimulus 'on' time (also 'response window'). (c) Fluorescence (calcium) trace from *Hydra* not stimulated (top) and stimulated with 25 psi (middle). Stimulation protocol in gray (bottom trace): 20 min no stimulation, 1 hr of repeated stimulation (1 s 'on,' 30 s 'off') and 20 min no stimulation. Stimulus 'on' times indicated with vertical lines. Magnification of 30 s fluorescence and stimulation protocol trace from one stimulation trial. (D) Raster plot of stimulus time-aligned spiking activity from multiple trials superimposed for *Hydra* not stimulated (top) and stimulated with 25 psi (bottom). (E) Mechanosensory response probability, fraction of trials (out of 119 total) that have at least one calcium spike (also contraction pulse) occurring during the 1 s response window (gray shaded region) when valve is pressurized. Large circles indicate average probability from all animals (N = 3) combined for each condition. Small circles indicate probability from a single animal. Significant pairwise comparisons are shown with brackets (one-way ANOVA with post-hoc Bonferroni correction). (F) Time interval between body contractions under each condition. Dashed line represents the time interval between stimuli (~31 s). Brackets indicate significant differences in a Kruskal–Wallis test with post-hoc Dunn–Sidak correction. (G) Percent of all body contractions that are a single pulse; brackets show significant pairwise comparisons from a one-way ANOVA with post-hoc Bonferroni correction. Error bars are standard error of mean (SEM); N = 3 *Hydra* for each condition; *$p<0.05$, **$p<0.01$, *****$p<0.00001$. (H) Table summarizing the mechanosensory response probability, time between contractions, and percent of contractions that are single pulses for each stimulus intensity (mean ± SEM or median ± SE). Source data for the quantitative characterization of mechanosensory response are available in Figure 2—source data 1.

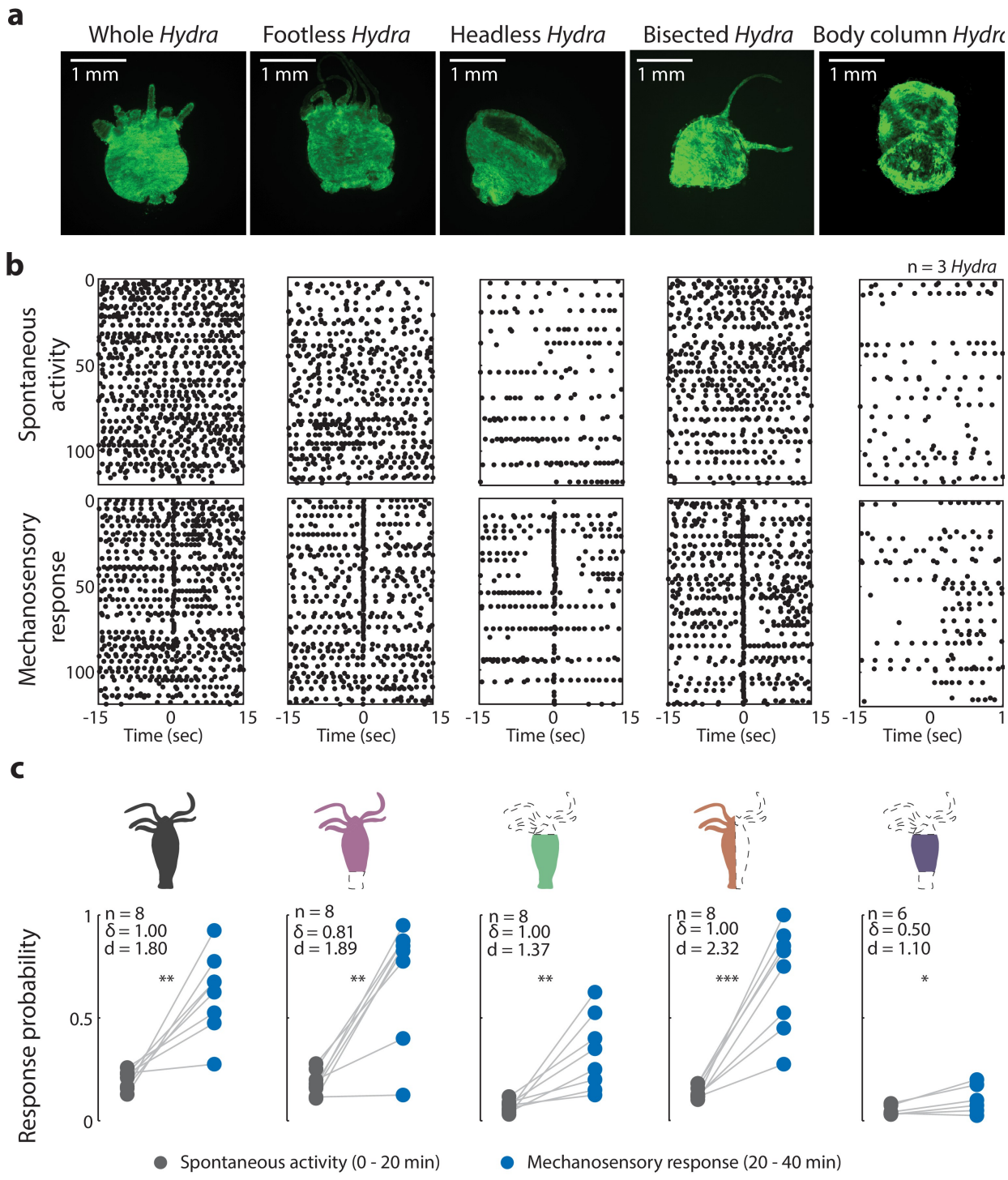


Figure C.3. Oral region is important for mechanosensory response. (A) Representative images of resection preparations (6–12 hr post-resections) of transgenic *Hydra* during contraction (GCaMP7b, endodermal epitheliomuscular cells): whole (or control), footless, headless, bisected, and body column animal. Entire frame region of interest (ROI) was used for analysis of the whole-body epithelial calcium activity. (B) Representative raster plot of stimulus time-aligned calcium spikes from three animals with multiple trials superimposed. Each black dot is a peak in calcium fluorescence identified as a contraction pulse. (C) Response probability,

fraction of trials that have at least one calcium spike (also contraction pulse) occurring within 1 s of stimulation onset. Gray dots are the mean contraction probability during no stimulation (t = 0–20 min) calculated from 1 s window shifted by ~0.3 s over 30 s intervals. Blue dots are contraction probability during stimulation (t = 20–40 min) calculated from 1 s response window during valve 'on.' Light gray lines connect the probabilities for spontaneous contraction and mechanosensory response for each individual. Cartoon schematics of *Hydra* indicate the resections performed. Excised body regions are outlined with a dashed line and unfilled area. Color-filled body regions indicate the portion of *Hydra* retained for the experiment. p-values from a paired t-test indicated as follows: n.s. = not significant; *p<0.05; **p<0.01; ***p<0.001. Source data for the mechanosensory response in resected animals are available in Figure 3—source data 1.

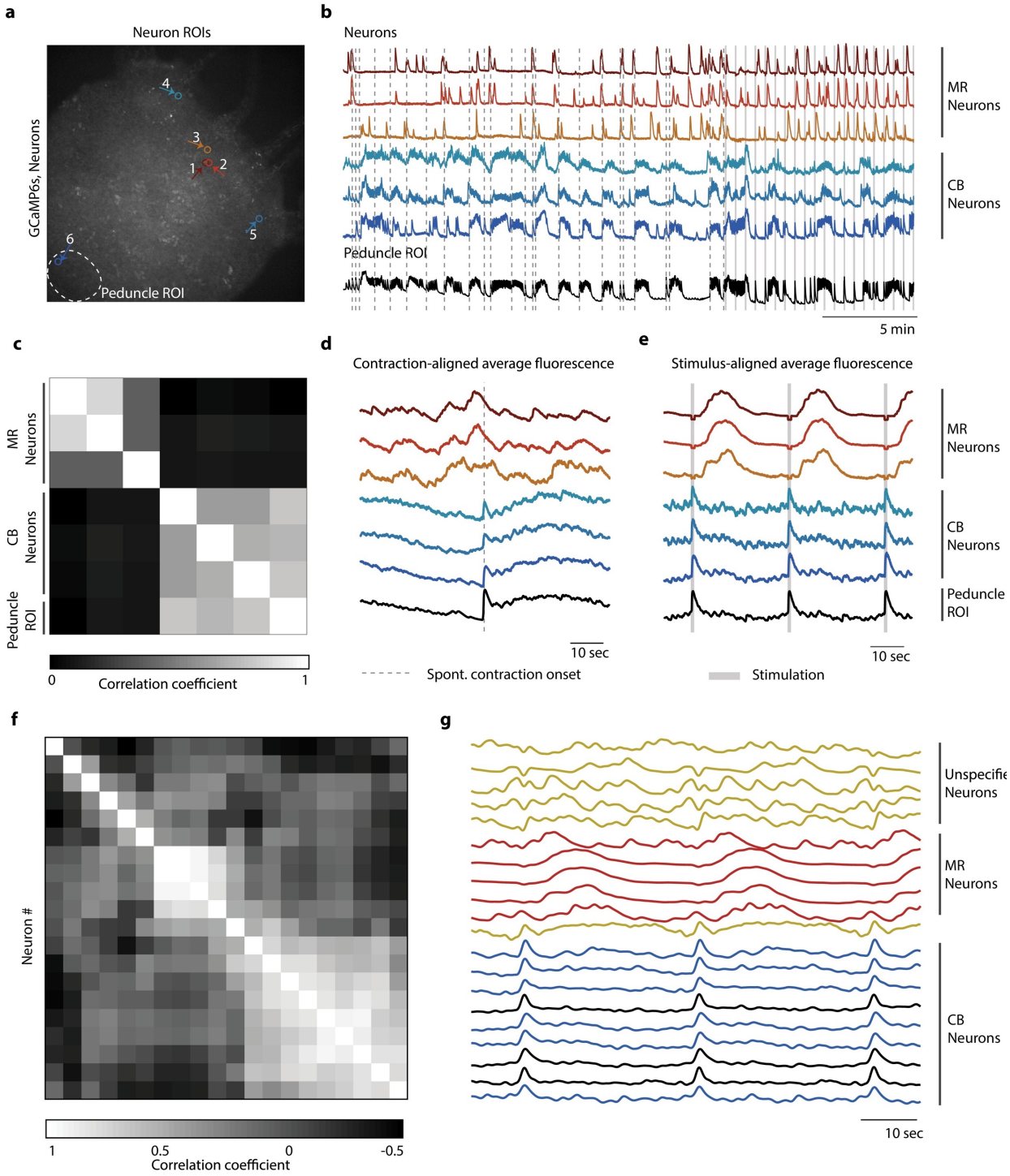


Figure C.4. Distinct networks of neurons involved in spontaneous and stimulated behaviors. (A) Fluorescent image of transgenic *Hydra* expressing GCaMP6s pan-neuronally. Individually tracked neurons regions of interest (ROIs) are indicated by arrows, and peduncle ROI is outlined with a white dashed circle. (B) Calcium fluorescence traces from single neurons (top six traces) and average calcium fluorescence from peduncle ROI (bottom trace). Mechanically responsive (MR) neurons are shown in shades of red. Contraction burst (CB) neurons are shown in shades of blue. (C) Heat map shows the correlation coefficients of

individually tracked neurons and peduncle ROI from *Hydra* 1, with color bar at the bottom. Correlation was computed using the entire 30 min calcium fluorescence from single neurons shown in (B). (D) Average calcium fluorescence traces from each of the neurons and peduncle ROI during spontaneous behaviors time-aligned with the onset of spontaneous body contractions. Dashed line indicates the onset of body contraction. (E) Average calcium fluorescence traces from each of the neurons and peduncle ROI during stimulated behaviors time-aligned with the onset of mechanical stimulation. Gray-shaded rectangle indicates mechanical stimulation. (F) Heat map shows the correlation coefficients of (G) the average calcium fluorescence traces from each of the neurons and peduncle ROI (traces in black) from three different *Hydra* during stimulated behaviors time-aligned with the onset of mechanical stimulation. Three groups of neurons were identified across the three *Hydra.* CB neurons (traces in blue), MR neurons (traces in red), and unspecified neurons (traces in yellow). (A-E) Calcium fluorescence traces and correlation analysis (entire 30 min of activity, 20 min of spontaneous activity, and 10 min of stimulated activity) from one representative *Hydra*. (F, G) Correlation analysis of the average stimulus-aligned calcium fluorescence (30 s stimulation interval average from 10 min of stimulated activity) of each neuron pooled multiple *Hydra* (N = 3 *Hydra*) to highlight the three groups of neurons identified: CB, MR, and unspecified. Source data for fluorescent calcium activity from single neurons are available in Figure 4—source data 1.

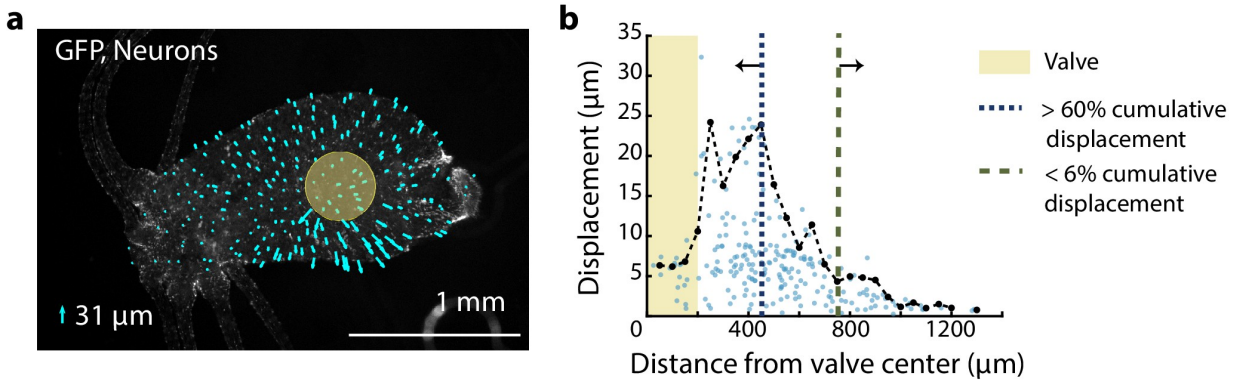


Figure C.5. Distribution of mechanical forces. (A) Fluorescent image (grayscale) of transgenic *Hydra* (nGreen) expressing GFP in neurons and neural progenitors. The oral region is on the left. Aboral/peduncle region is on the right. Arrows (cyan) overlayed indicate the displacement of individual cells immediately before and after the valve is pressurized for stimulation (~50 ms). The lengths of the arrows are enlarged to be visible with the scaled size of the largest displacement ~31 µm as indicated by the legend (lower left). Semi-transparent (yellow) circle indicates the location of the valve (400 µm diameter). (B) Displacement of neurons due to stimulation varies with the location of the neurons from the valve center. Each semi-transparent blue dot represents a single neuron (n = 222 neurons). Black dots (connected by black dashed line) represent the average displacement of the highest three displacements within a 50 µm radial band. Microfluidic valve has a radius of 200 µm indicated by the yellow-shaded region. Majority of the cumulative displacement (>60%) is localized to a 250 µm radius bordering the valve indicated by the blue dashed line at 450 µm. Neurons more than 750 µm from the microfluidic valve center have negligible displacement (<6% cumulative displacement) indicated by the green dashed line.

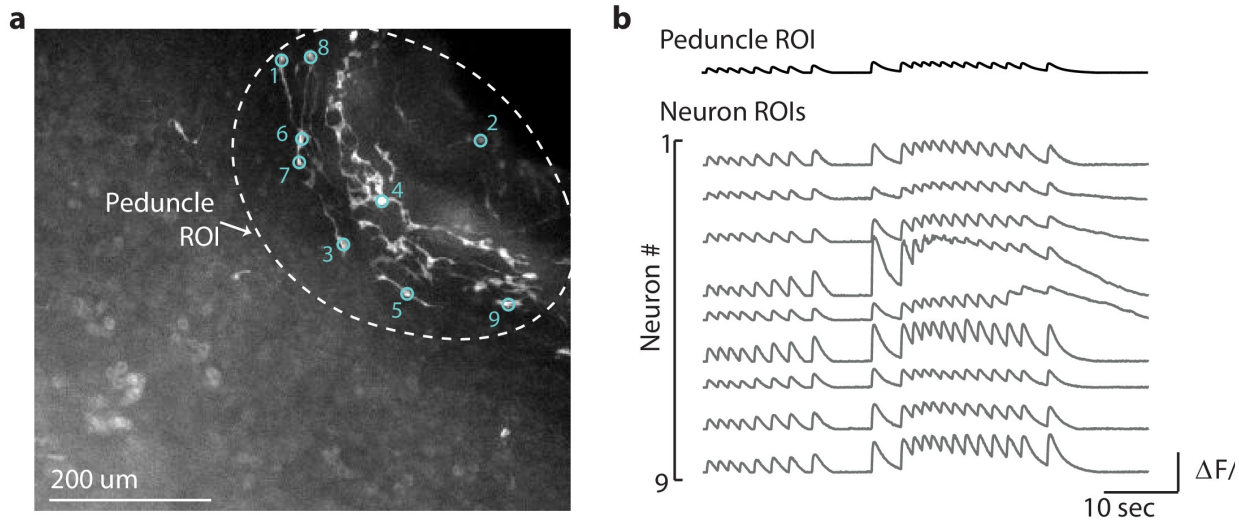


Figure C.6. Average calcium fluorescence from large peduncle region of interest (ROI) correlated with calcium fluorescence from smaller ROIs for individual peduncle neurons. (A) Fluorescent image of transgenic *Hydra* expressing GCaMP6s pan-neuronally. Dashed while circle indicates the peduncle ROI. Small cyan circles indicate the individual neuron ROIs. Peduncle region is in the top right. The oral end of the *Hydra* (not shown in the image) is oriented towards the bottom-left corner. (B) Average calcium fluorescence trace (black) from the large peduncle ROI. Average calcium traces (gray) from individual neurons in the peduncle.

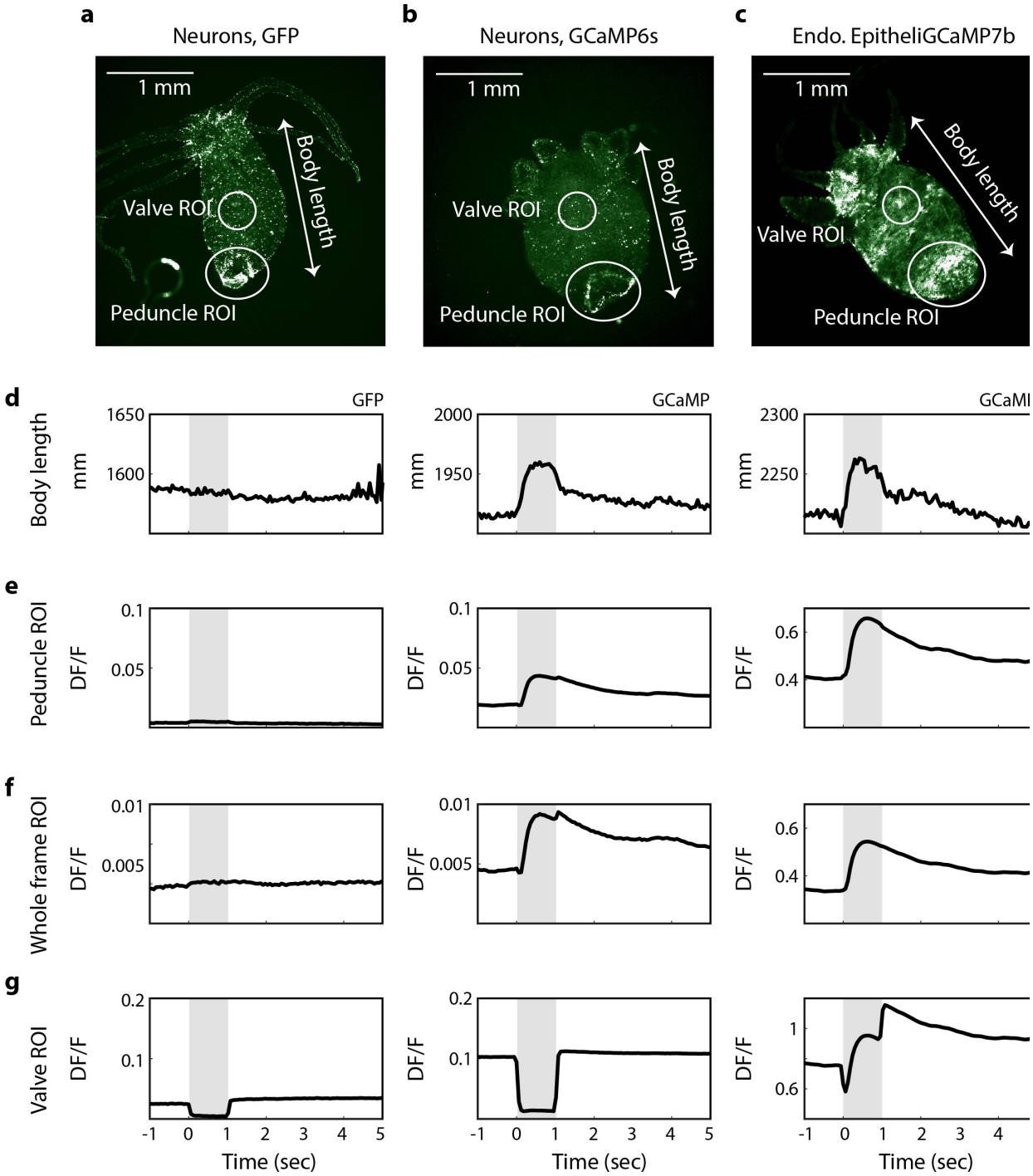


Figure C.7. Spikes in calcium fluorescence are due to calcium activity not motion artifacts. (A) Fluorescence image of transgenic *Hydra* (nGreen) expressing GFP in neurons false colored with hot green colormap. Neurons appear in white. (B) Fluorescence image of transgenic *Hydra* expressing GCaMP6s in neurons false colored with hot green colormap. (C) Fluorescence image of transgenic *Hydra* expressing GCaMP7b in endodermal epitheliomuscular cells false colored with hot green colormap. (A-C) Body length (white arrow), valve region of interest (ROI) (top circles), peduncle ROI (larger bottom circles), and whole-frame ROI (entire frame square) annotated on all three images for comparison. (D) Average

body length trace time-aligned to stimulus onset from GFP *Hydra* (left) and GCaMP *Hydra* (middle, right). Light gray-shaded region indicates when the stimulation is on. (E) Average fluorescence trace calculated from the annotated peduncle ROI from GFP *Hydra*(left) and GCaMP *Hydra* (middle, right). (F) Average fluorescence trace calculated from the whole-frame ROI from GFP *Hydra* (left) and GCaMP *Hydra* (middle, right). (G) Average fluorescence trace calculated from the annotated valve ROI from GFP *Hydra* (left) and GCaMP *Hydra* (middle, right).

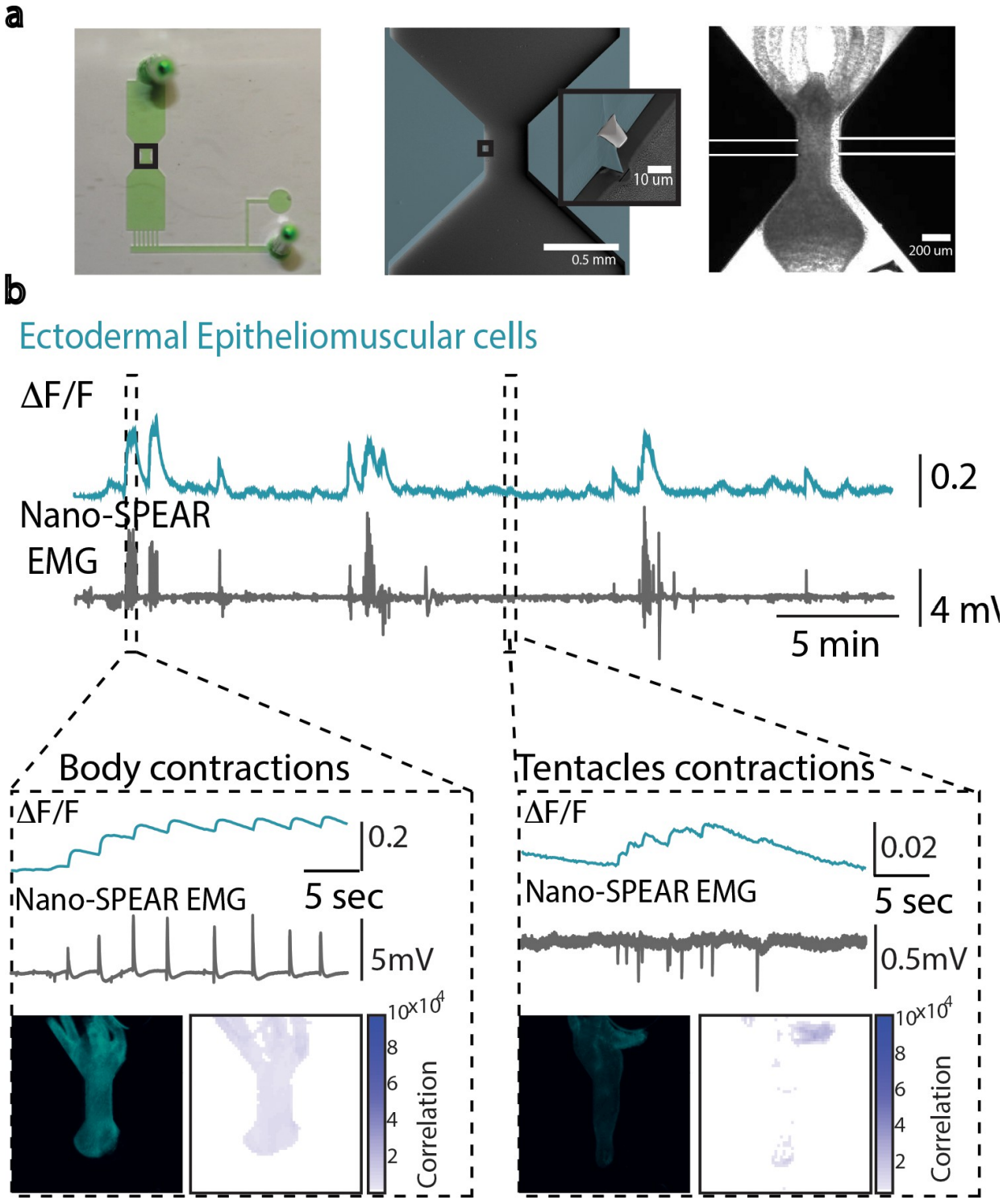


Figure C.8. Simultaneous electrophysiology and calcium imaging of ectodermal epitheliomuscular cells. (A) Photograph (left) of a microfluidic immobilization chamber filled with green dye. Black box highlights the recording region in the microfluidic chamber (110 µm tall). False-colored scanning electron micrograph (middle) shows the recording region (blue, photoresist; light gray, Pt; dark gray, silica) on the nano-SPEAR chip (50 µm tall). Inset shows a zoom-in of the Pt electrode (light gray) suspended midway between the top and bottom of the

photoresist sidewall (blue). Brightfield image (right) shows *Hydra* immobilized in the microfluidic chamber placed on top of the nano-SPEAR chip with combined 160 μm tall recording region. (B) Simultaneous electrophysiology and calcium imaging in transgenic *Hydra* (GCaMP6s, ectodermal epitheliomuscular cells). Top trace shows mean fluorescence (ΔF/F) from ectodermal epitheliomuscular cells (whole-frame region of interest [ROI]) in *Hydra* (GCaMP, ectodermal) (10 Hz). Bottom trace shows simultaneously recorded electrical activity from the *Hydra* (1 kHz). High- and low-activity periods identified based on peak amplitude. Inset: left box shows correlation during high-activity period, contraction bursts. Traces show peaks in fluorescence (top trace) coinciding with peaks in electrophysiology (bottom trace). A representative fluorescence image shows high levels of fluorescence thus calcium activity in the entire body during contraction burst events. Correlation map spatially plotting the correlation coefficient shows the entire body has calcium activity correlated with electrophysiology. Inset: right box shows correlation during low-activity period, tentacle pulses. Traces show peaks in fluorescence (top trace) coinciding with peaks in electrophysiology (bottom trace). A representative fluorescence image shows high levels of fluorescence in the tentacle region during tentacle pulses. Cross-correlation map shows the tentacle region has calcium activity correlated with electrophysiology. Almost all spikes in electrophysiology coincide with spike in fluorescence from ectodermal epitheliomuscular cells (and behavioral contractions).

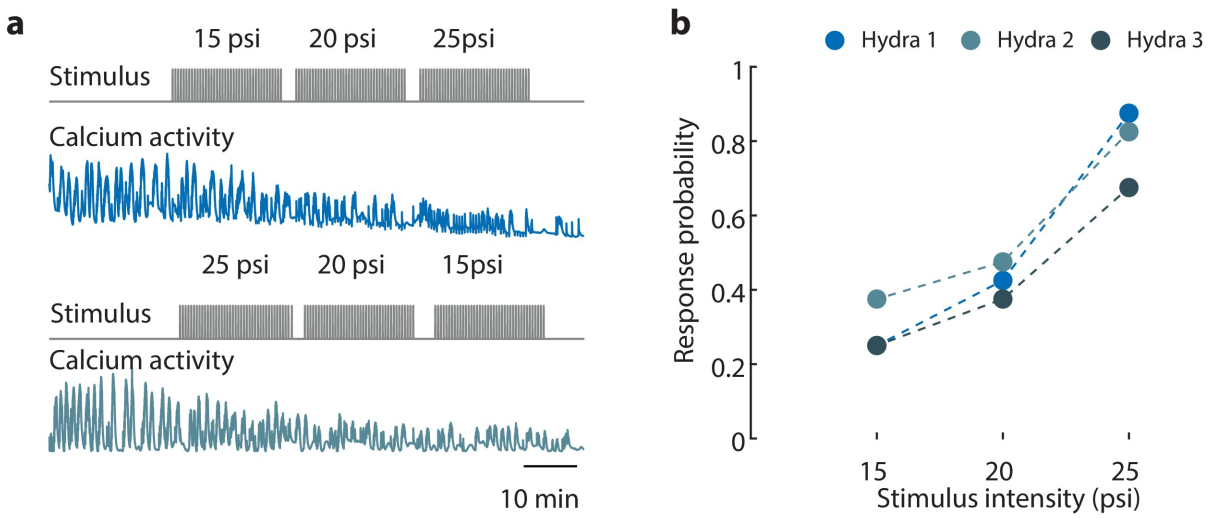


Figure C.9. *Hydra*'s epitheliomuscular response is dependent on the mechanical stimulus intensity. (A) Gray trace (top) is the stimulation protocol. 20 min no stimulation, 20 min of repeated stimulation (1 s 'on,' 30 s 'off') at 15 psi, 20 min of repeated stimulation (1 s 'on,' 30 s 'off') at 20 psi, and 20 min of repeated stimulation (1 s 'on,' 30 s 'off') at 25 psi. Stimulus 'on' times indicated by vertical lines. Entire frame region of interest (ROI) used for analysis of the whole-body epithelial calcium activity. Representative calcium fluorescence trace (blue) from endodermal epitheliomuscles (GCaMP7b) from animal stimulated with three different stimulus intensities for 20 min each (15, 20, and 25 psi). Representative calcium fluorescence trace (teal) from endodermal epitheliomuscles (GCaMP7b) from animal stimulated with three different stimulus intensities for 20 min each (25, 20, and 15 psi). The decrease in fluorescence amplitude observed for both increasing in and decreasing stimulus intensity is due to photobleaching of the calcium indicator. (B) Mechanosensory response probability at different stimulus intensities (N = 3 animals). Response probability, fraction of trials that have at least one calcium spike (also contraction pulse) occurring within 1 s of stimulation onset.

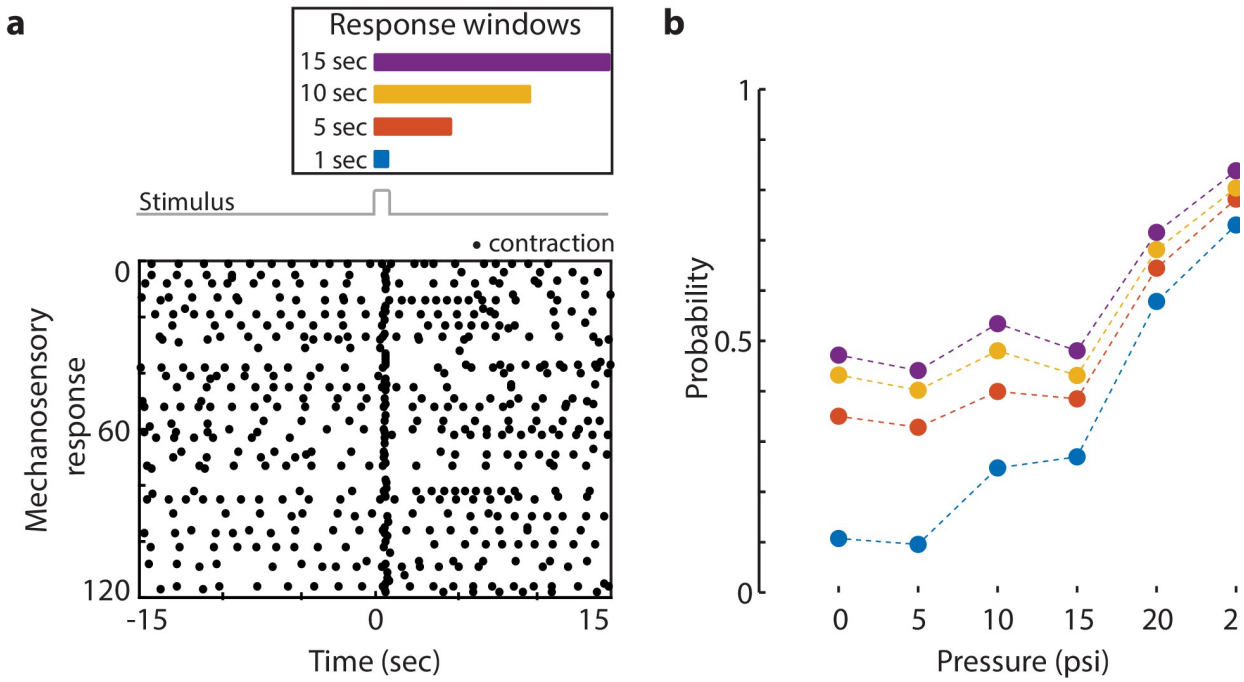


Figure C.10. Mechanosensory response window. (A) Different response windows tested (1 s in blue; 5 s in orange; 10 s in yellow; and 15 s in purple). (Top) The lengths of the rectangles correspond to time (s) in the raster plot below. Stimulus trace in gray. Representative raster plot of time-aligned contraction pulses from multiple trials superimposed from one animal stimulated at 20 psi every 31 s for 60 min. (Bottom) Each black dot is a spike in calcium fluorescence identified as a contraction pulse. Stimulus is applied from 0 to 1 s as indicated by a step in stimulus trace. (B) Response probability, fraction of trials that have at least one calcium spike (also contraction pulse) occurring during different response windows (1 s in blue; 5 s in orange; 10 s in yellow; and 15 s in purple within stimulation onset).

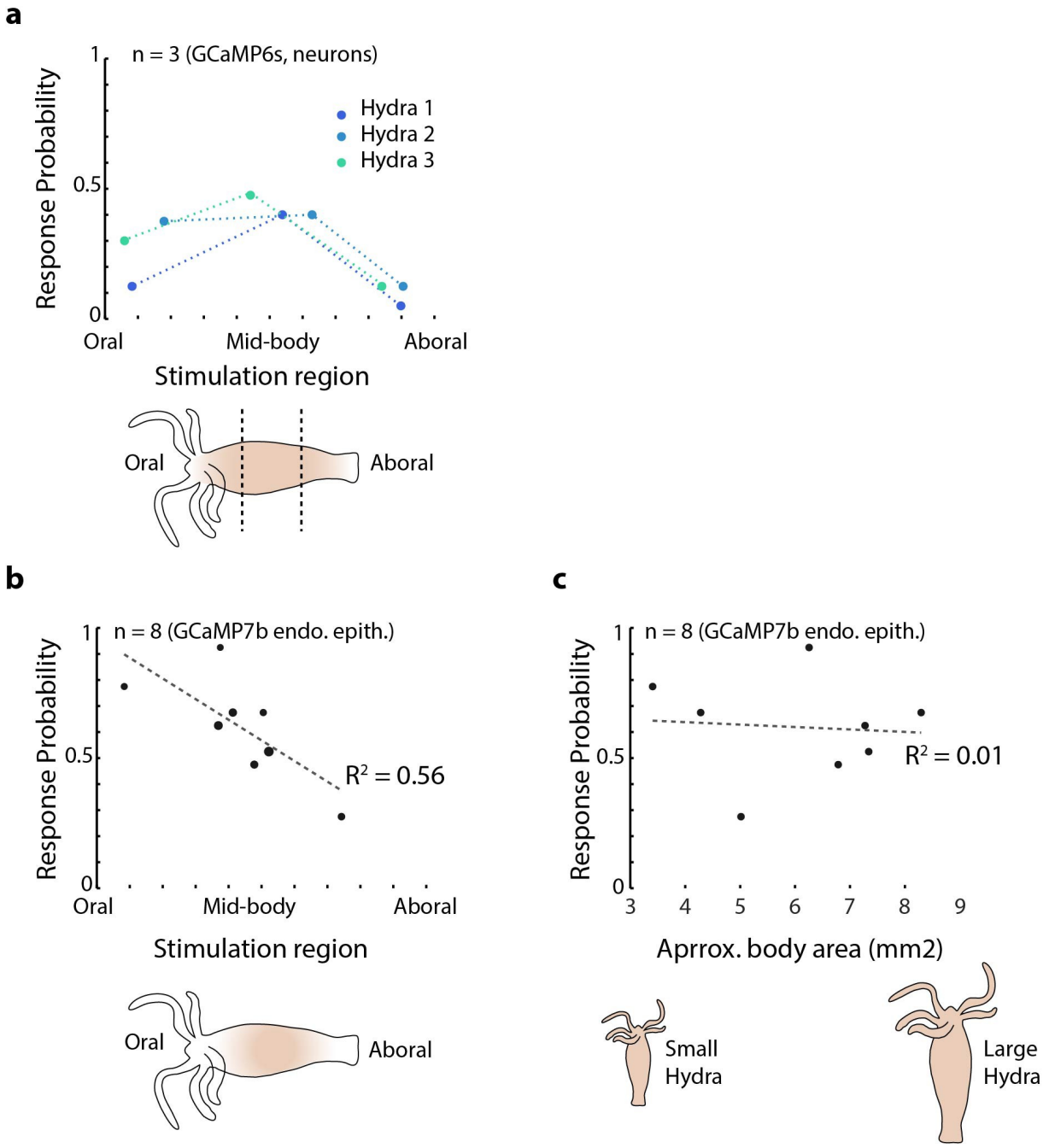


Figure C.11. Mechanical sensitivity of different body regions in *Hydra*. (A) Response probability of transgenic *Hydra* (N = 3 animals expressing GCaMP6s in neurons) stimulated at three different body regions: oral, mid-body, and aboral. Annotated *Hydra* below the plot indicates the three stimulation regions used. (B) Response probability of transgenic *Hydra* (n = 8 animal expressing GCaMP7b in endodermal epitheliomuscular cells). Response probability is calculated using average calcium fluorescence from neurons in the peduncle region of interest (ROI) in (A) and using average calcium fluorescence from the endodermal epitheliomuscular cells from the entire body in (B) and (C).

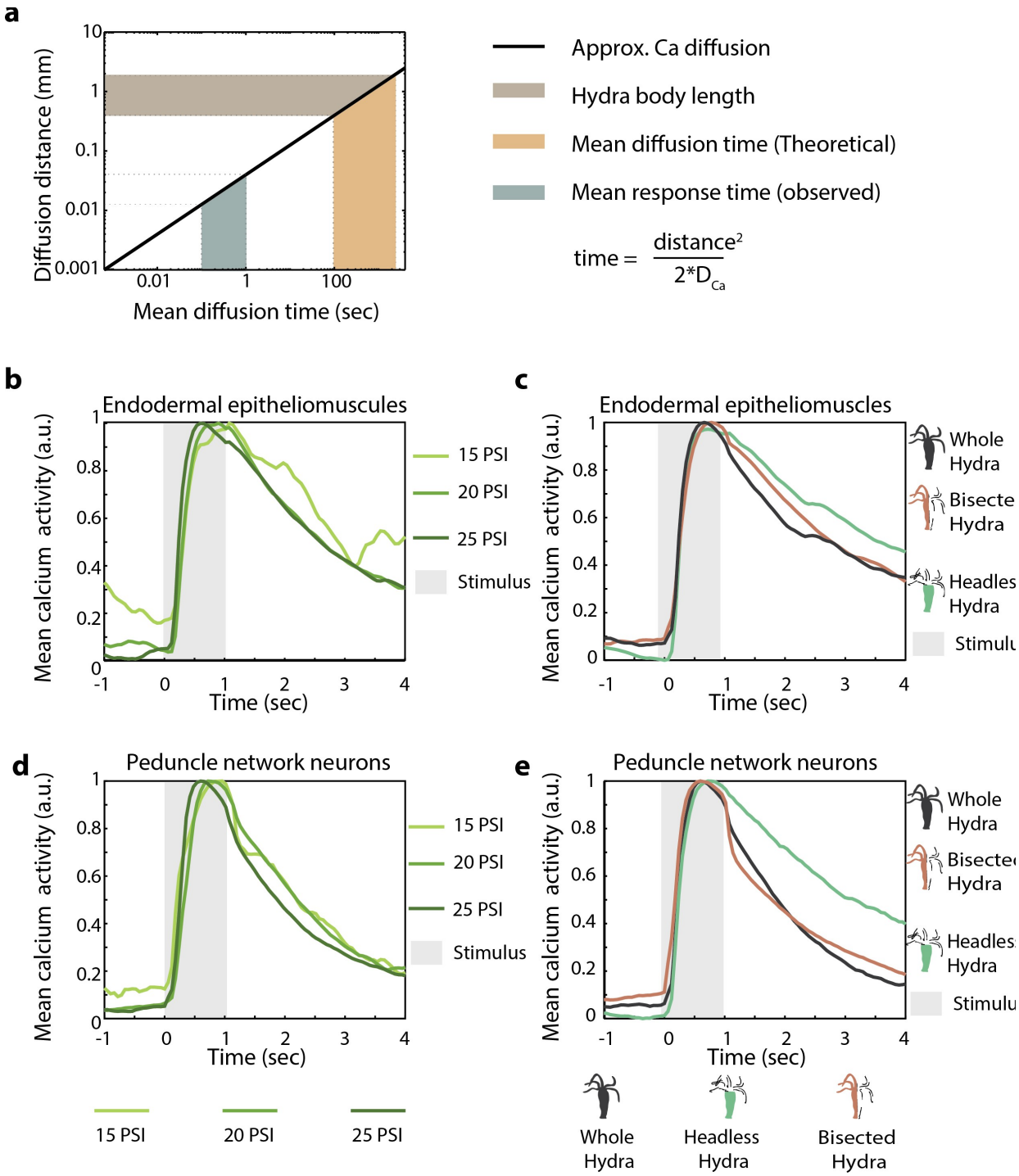


Figure C.12. *Hydra's* mechanosensory response time is faster than passive calcium diffusion through epitheliomuscular cells. (A) Distance calcium diffuses passively over given time (left, black line plot) approximated using passive diffusion equation (right). Light brown-shaded region indicates the range of body length (~0.5–1 mm) of *Hydra* in microfluidic chambers used for the experiments. Teal-shaded region indicates the average time between stimulus onset and observed spike in fluorescence (mean response time 0.5–1 s). Yellow-shaded region is the theoretical mean calcium diffusion time calculated assuming passive diffusion. (B-E) Mean calcium fluorescence time-aligned to stimulus and normalized. Stimulus is applied at

0 s (valve ‘on’ from 0 to 1 s). (B) Mean calcium activity from the peduncle network of neurons at different stimulus intensities. (C) Mean calcium activity from endodermal epitheliomuscular cells at different stimulus intensities. (B,C) The line colors correspond to different stimulus intensities shown at the bottom. (D) Mean calcium activity from peduncle network of neurons in different body resections stimulated at 20 psi (pressure for ~60% response probability in whole animals). (E) Mean calcium activity from endodermal epitheliomuscular cells in different body resections different body resections stimulated at 22 psi (pressure for ~60% response probability in whole animals). (D,E) The line colors correspond to the different resections shown at the bottom with cartoon schematic of the resected *Hydra* with the same color.

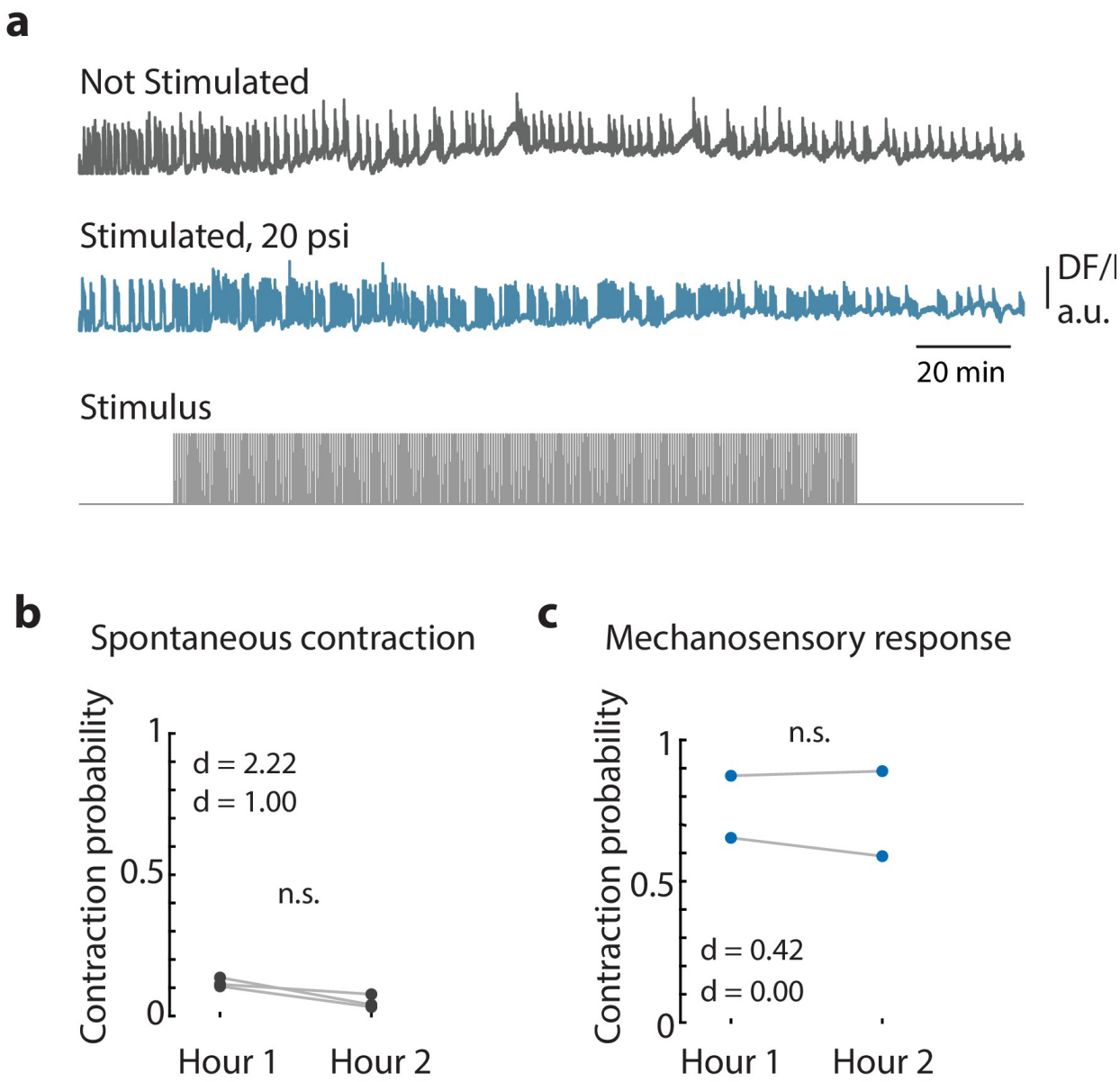


Figure C.13. Long-term mechanical stimulation. (A) Representative calcium fluorescence trace from the peduncle region from an animal (GcaMP6s, neurons) not stimulated and animal mechanically stimulated with 20 psi. Stimulus protocol in gray, 20 min no stimulation, 120 min of repeated stimulation (1 s 'on,' 30 s 'off') at 20 psi followed by no stimulation. (B) Spontaneous and (C) stimulated (mechanosensory response) contraction pulse probabilities for the first and second hour of stimulation compared for each animal. Gray circles are mean contraction pulse probability during no stimulation calculated from an average of 1 s window shifted by ~0.3 s over 30 s intervals (N = 3). Blue circles are contraction probability during stimulation calculated from 1 s response window during valve 'on' (N = 2). Light gray lines indicate the change in probabilities from hour 1 to hour 2 for each individual (paired t-test, n.s. = not significant).

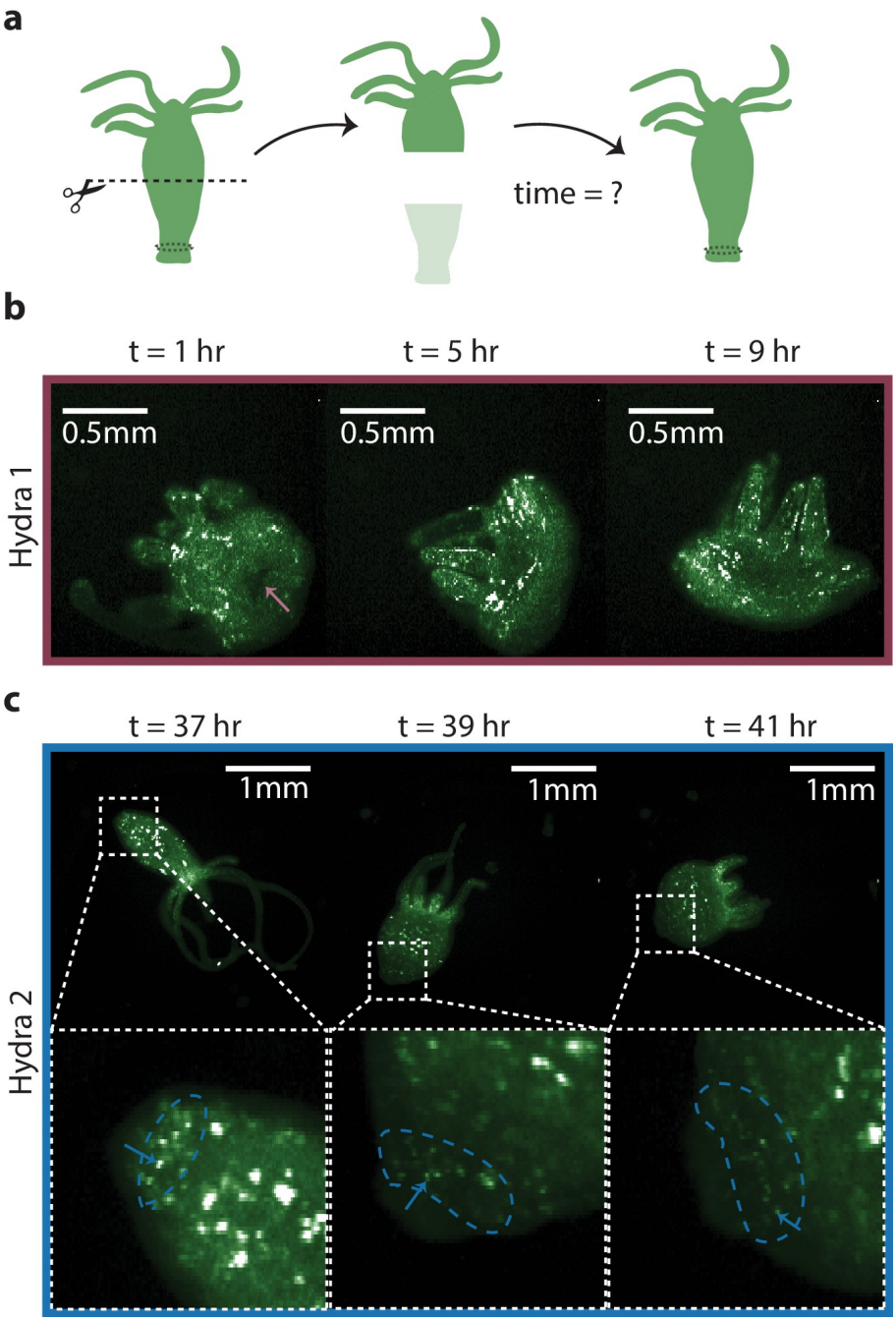


Figure C.14. Regeneration of the peduncle network. (A) Summary schematic of the experiment performed. Dashed line indicates where the incision was made to remove the lower half of the transgenic *Hydra* body (GcaMP6s, neurons) to discover when the peduncle neuronal network is regenerated. (B) Fluorescence images of 'footless' *Hydra* during body contractions at various timepoints t = 1 hr, 5 hr, and 9 hr after bisection. Pink arrow indicates the open wound visible at t = 1 hr but not at other timepoints. (C) Fluorescence images of another 'footless' *Hydra* during body contractions at various timepoints t = 37 hr, 39 hr, and 41 hr. Blue dashed circle indicates the region where the peduncle network is found. Blue arrows indicate some of the neurons in the peduncle network with clearly well-connected neurons at t = 41 hr.

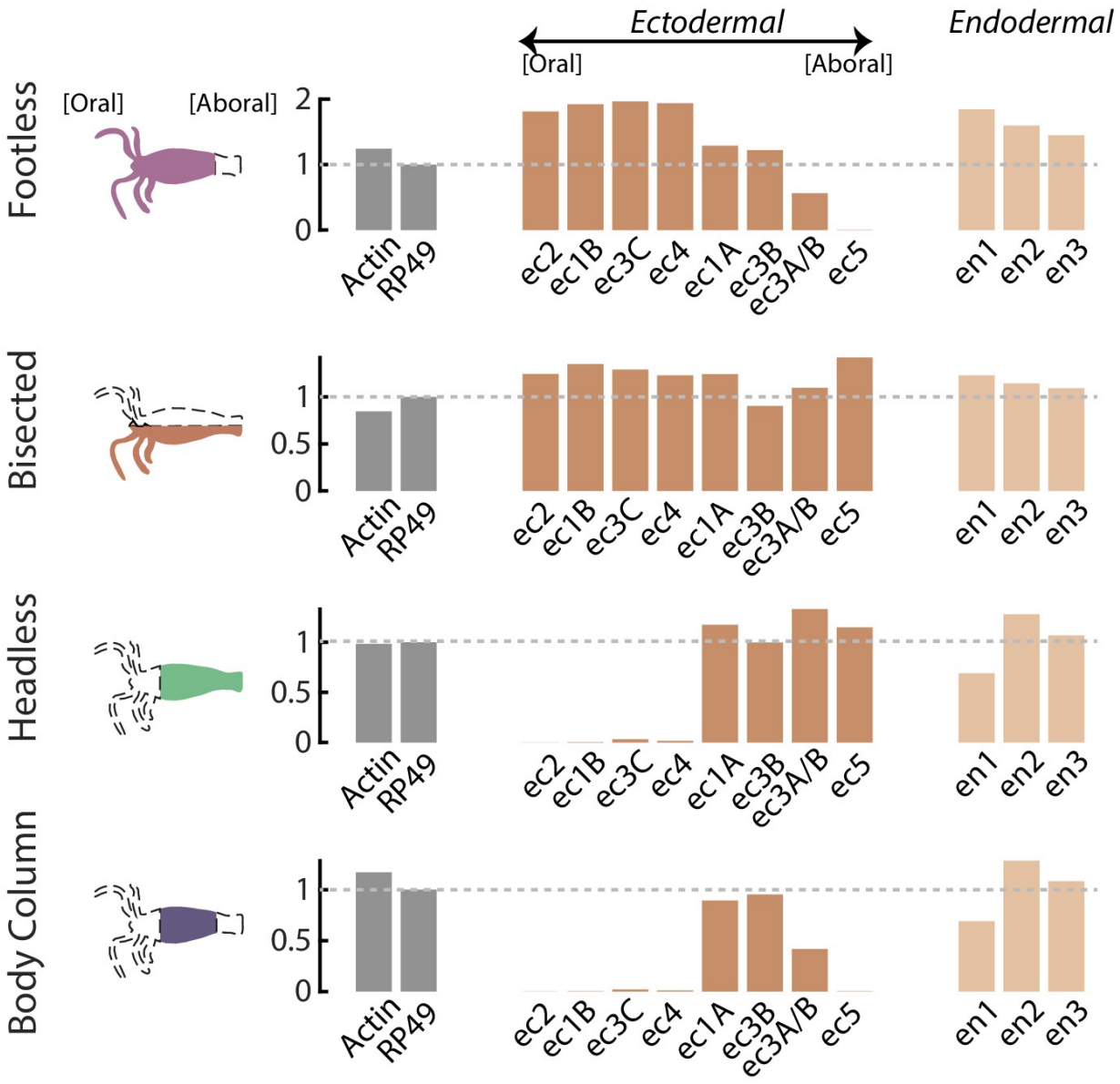


Figure C.15. RT-qPCR analyses of neuron subtype-specific gene expression in resected animals demonstrates loss of specific neuron subtypes. RT-qPCR was used to test for the loss of specific neuron subtypes in whole, tube (no head or foot), headless, footless, and bisected *Hydra* using uniquely expressed biomarkers for each subtype. There is not a specific biomarker for ec3A (located in the basal disk), so the marker used to test for the presence or absence of this cell type is also expressed in ec3B (located in the body column). Therefore, expression of the ec3A/B marker gene is reduced, but not completely lost in animals with resected feet ('tube' and 'footless'). However, ec5 expression (located in the peduncle above the basal disk) is completely lost in animals with resected feet, thus it is clear that the ec3A subtype is completely lost in these animals. Biomarkers for neuron subtypes located in the head and tentacles (ec1B, ec2, ec3C, ec4) are completely lost in animals with resected heads ('headless' and 'tube'). Data were analyzed with the $2^{\wedge-\Delta\Delta Ct}$ method, and results were normalized to the housekeeping gene *RP49* and to expression in whole animals.

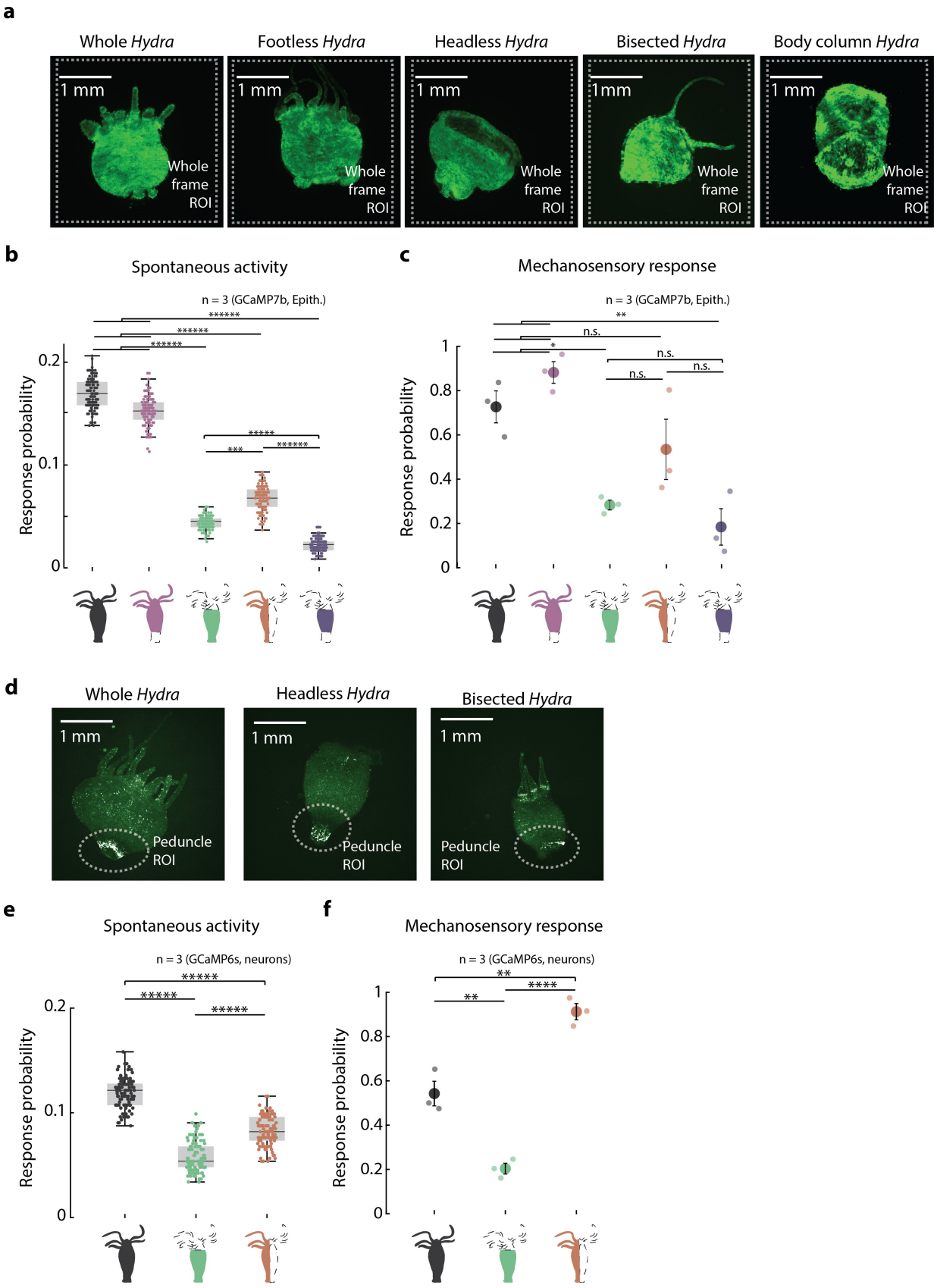

a
Whole Hydra
Footless Hydra
Headless Hydra
Bisected Hydra
Body column Hydra
1 mm
Whole frame ROI
b
Spontaneous activity
n = 3 (GCaMP7b, Epith.)
Response probability
c
Mechanosensory response
n = 3 (GCaMP7b, Epith.)
Response probability
n.s.
d
Whole Hydra
Headless Hydra
Bisected Hydra
Peduncle ROI
e
Spontaneous activity
n = 3 (GCaMP6s, neurons)
Response probability
f
Mechanosensory response
n = 3 (GCaMP6s, neurons)
Response probability

Figure C.16. Mechanosensory response from endodermal epitheliomuscular cells and neurons in resected Hydra. (A) Representative images of resection preparations of transgenic *Hydra* (GCaMP7b, endodermal epitheliomuscular cells): whole (or control), 'footless,' 'headless,' 'bisected,' and 'body column.' White dashed square indicates the whole-frame region of interest (ROI) used for quantifying response. (B) Spontaneous probability of at least one spiking event (or contraction pulse) occurring during a random 1 s window (1 s window slide by ~0.3 s across 30 s stimulation intervals; Kruskal–Wallis test with post-hoc Dunn–Sidak correction). (C) Mechanosensory response probability during 1 s of valve on. Response probability, fraction of trials that have at least one calcium spike (also contraction pulse) occurring within 1 s of stimulation onset. Large circles indicate average values from all animals combined for each condition. Small circles indicate probability from individual animals (one-way ANOVA with post-hoc Bonferroni correction). Error bars are standard error of mean (SEM) (N = 3 *Hydra* for each condition; analysis from 60 min of stimulation, which is imaging t = 20–80 min; n.s. = not significant, **p<0.01, ****p<0.0001, *****p<0.00001). (D) Representative images of resection preparations of transgenic *Hydra* (GCaMP6s, neurons): whole (or control), 'headless,' and 'bisected.' White dashed circle indicates the peduncle ROI used for quantifying response. (E) Spontaneous probability of at least one spiking event (or contraction pulse) occurring during a random 1 s window (1 s window slide by ~0.3 s across 30 s stimulation intervals; Kruskal–Wallis test with post-hoc Dunn–Sidak correction). (F) Mechanosensory response probability during 1 s of valve on. Large circles indicate average values from all animals combined for each condition. Response probability, fraction of trials that have at least one calcium spike (also contraction pulse) occurring within 1 s of stimulation onset. Small circles indicate probability from individual animals (one-way ANOVA with post-hoc Bonferroni correction). Error bars are standard error of mean (SEM) (N = 3 *Hydra* for each condition; analysis from 60 min of stimulation, which is imaging t = 20–80 min; n.s. = not significant, **p<0.01, ****p<0.0001, *****p<0.00001).

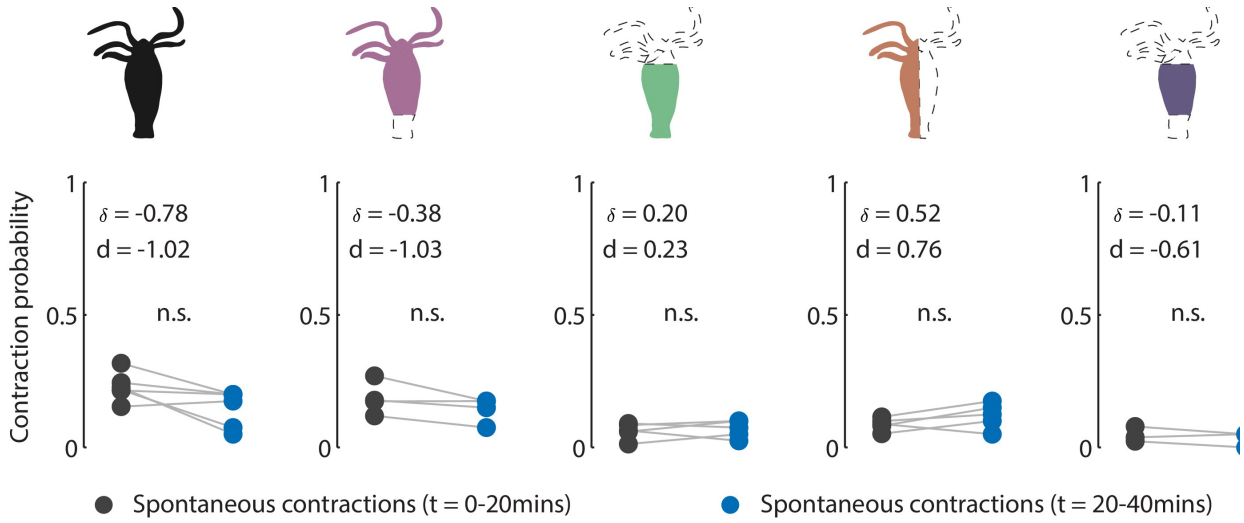


Figure C.17. Contraction activity in non-stimulated animals. Cartoon schematic of resection preparations of transgenic *Hydra* (GCaMP7b, endodermal epitheliomuscular cells): whole (or control), 'footless,' 'headless,' 'bisected,' and 'body column' animals. Entire frame region of interest (ROI) used for analysis of the whole-body epithelial calcium activity. Gray circles are mean contraction probability during t = 0–20 min. Blue circles are mean contraction probability during t = 20–40 min (this time corresponds to when stimulated animals receive mechanical stimuli). Mean probability calculated from 1 s window shifted by ~0.3 s over 30 s intervals. Response probability, fraction of trials that have at least one calcium spike (also contraction pulse) occurring within 1 s of stimulation onset. Pressure in valves = 0 psi. Light gray lines pair the spontaneous contractions probability from the first 20 min interval with the second 20 min interval from each individual animal. Whole N = 4, footless N = 3, headless N = 3, bisected N = 5, body column N = 3 *Hydra*, paired t-test, n.s. = not significant.

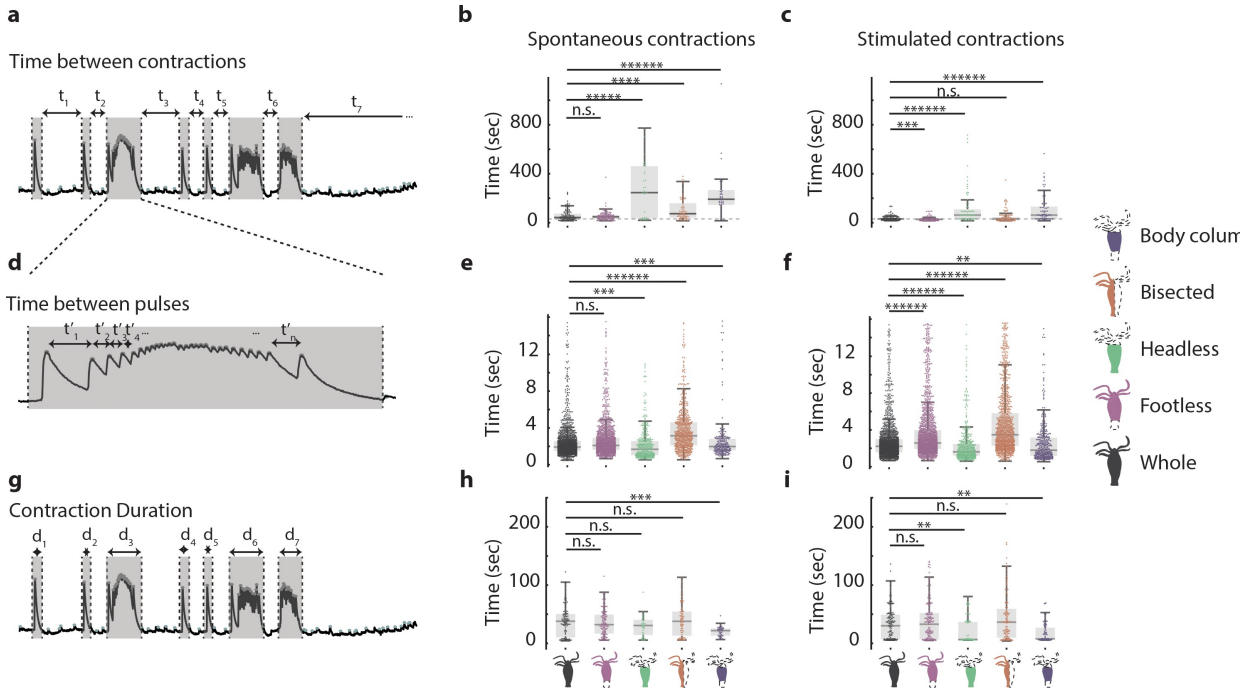


Figure C.18. Hypostome and peduncle nerve rings work together to coordinate contractile behavior. (A) Representative fluorescence trace used to calculate time interval between contractions. (B) Time interval between spontaneous contractions in animals with different resections: whole, 'footless,' 'headless,' 'bisected,' and 'body column.' (C) Time interval between stimulated contractions in animals with different resections. (D) Representative fluorescence trace used to calculate time interval between contraction pulses. For illustration purpose, only a select few of the time intervals are shown. (E) Time interval between spontaneous contraction pulses in animals with different resections. (F) Time interval between stimulated contraction pulses in animals with different resections. (G) Representative fluorescence trace used to calculate contraction duration. (H) Duration of spontaneous contractions in animals with different resections. (I) Duration of stimulated contractions in animals with different resections. N = 3 *Hydra* (GcaMP7b, ectodermal epitheliomuscles) per resection Kruskal–Wallis test with Dunn–Sidak correction, n.s. = not significant, **p<0.01, ***p<0.001, ****p<0.0001, *****p<0.00001, ******p<0.000001.

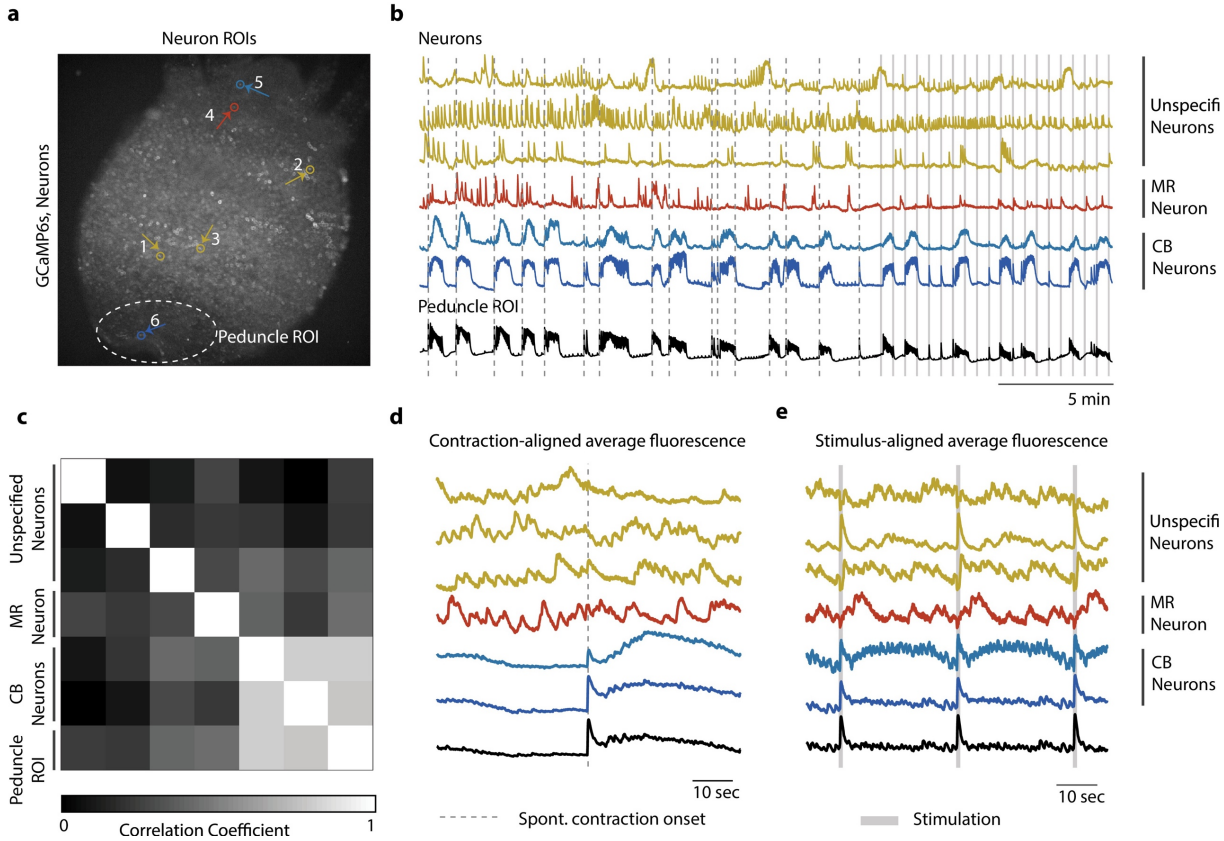


Figure C.19. Single-cell correlation analysis. (A) Fluorescence image of transgenic *Hydra* (#2) expressing GCaMP6s pan-neuronally. Individually tracked neuronal regions of interest (ROIs) are indicated with circles. The neurons are numbered in the order their traces appear in (C–E) with the same colors. Peduncle ROI outlined with white dashed line. (B) Calcium fluorescence traces from single neurons (top six traces) and average calcium fluorescence from peduncle ROI (bottom trace). Mechanically responsive (MR) neurons are shown in shades of red. Contraction burst (CB) neurons are shown in shades of blue. Unspecified neurons are shown in yellow, and their activity do not resemble any of the previously identified neuronal networks (contraction burst, rhythmic potential, or the mechanically responsive reported here). (C) Heat map shows the correlation coefficients of individually tracked neurons and peduncle ROI. Color bar is at the bottom. (D) Average calcium fluorescence traces from each of the neurons and peduncle ROI during spontaneous behaviors time-aligned with the onset of spontaneous body contractions. Dashed line indicates the onset of body contraction. (E) Average calcium fluorescence traces from each of the neurons and peduncle ROI during stimulated behaviors time-aligned with the onset of mechanical stimulation. Gray-shaded rectangle indicates mechanical stimulation.

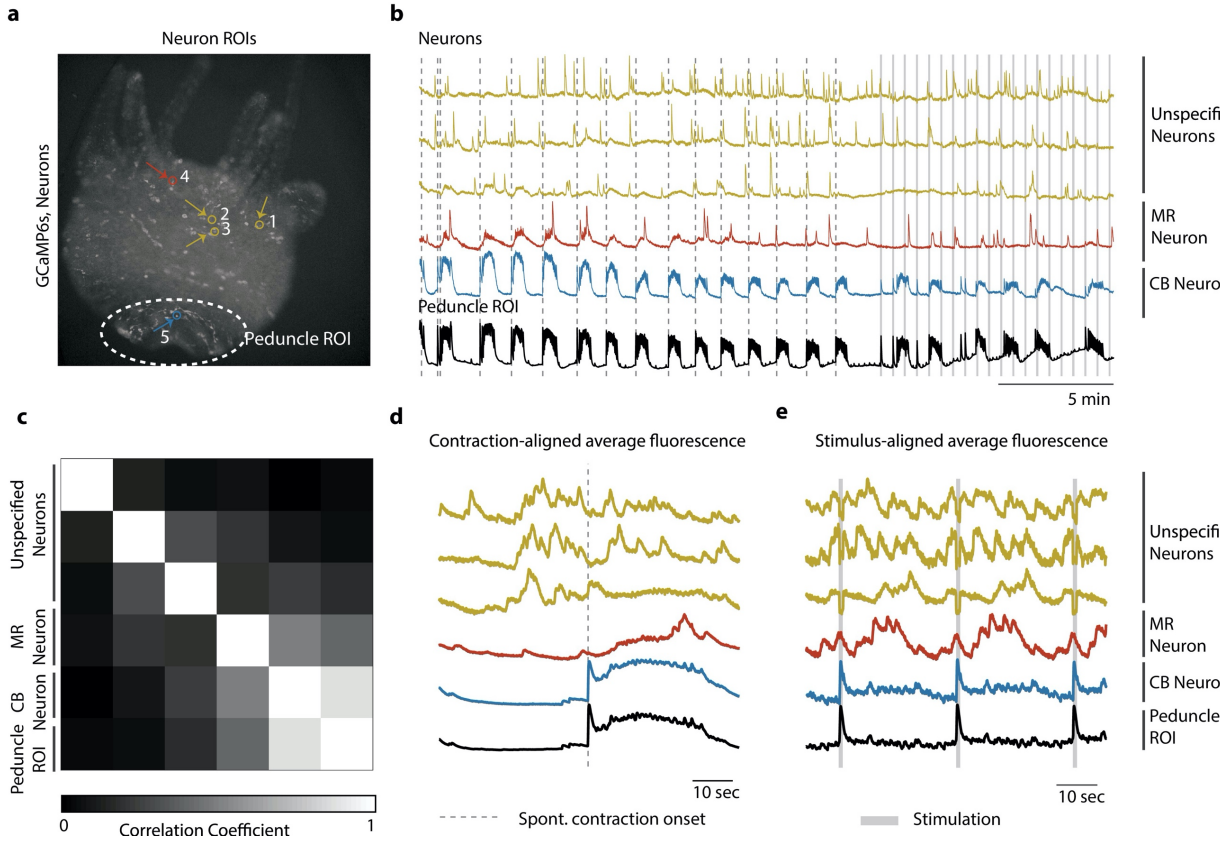


Figure C.20. Single-cell correlation analysis. (A) Fluorescence image of transgenic *Hydra* (#3) expressing GCaMP6s pan-neuronally. Individually tracked neuronal regions of interest (ROIs) are indicated with circles. The neurons are numbered in the order their traces appear in (C–E) with the same colors. Peduncle ROI outlined with white dashed line. (B) Calcium fluorescence traces from single neurons (top five traces) and average calcium fluorescence from peduncle ROI (bottom trace). Mechanically responsive (MR) neurons are shown in shades of red. Contraction burst (CB) neurons are shown in shades of blue. Unspecified neurons are shown in yellow, and their activity do not resemble any of the previously identified neuronal networks (contraction burst, rhythmic potential, or the mechanically responsive reported here). (C) Heat map shows the correlation coefficients of individually tracked neurons and peduncle ROI. Color bar is at the bottom. (D) Average calcium fluorescence traces from each of the neurons and peduncle ROI during spontaneous behaviors time-aligned with the onset of spontaneous body contractions. Dashed line indicates the onset of body contraction. (E) Average calcium fluorescence traces from each of the neurons and peduncle ROI during stimulated behaviors time-aligned with the onset of mechanical stimulation. Gray-shaded rectangle indicates mechanical stimulation.

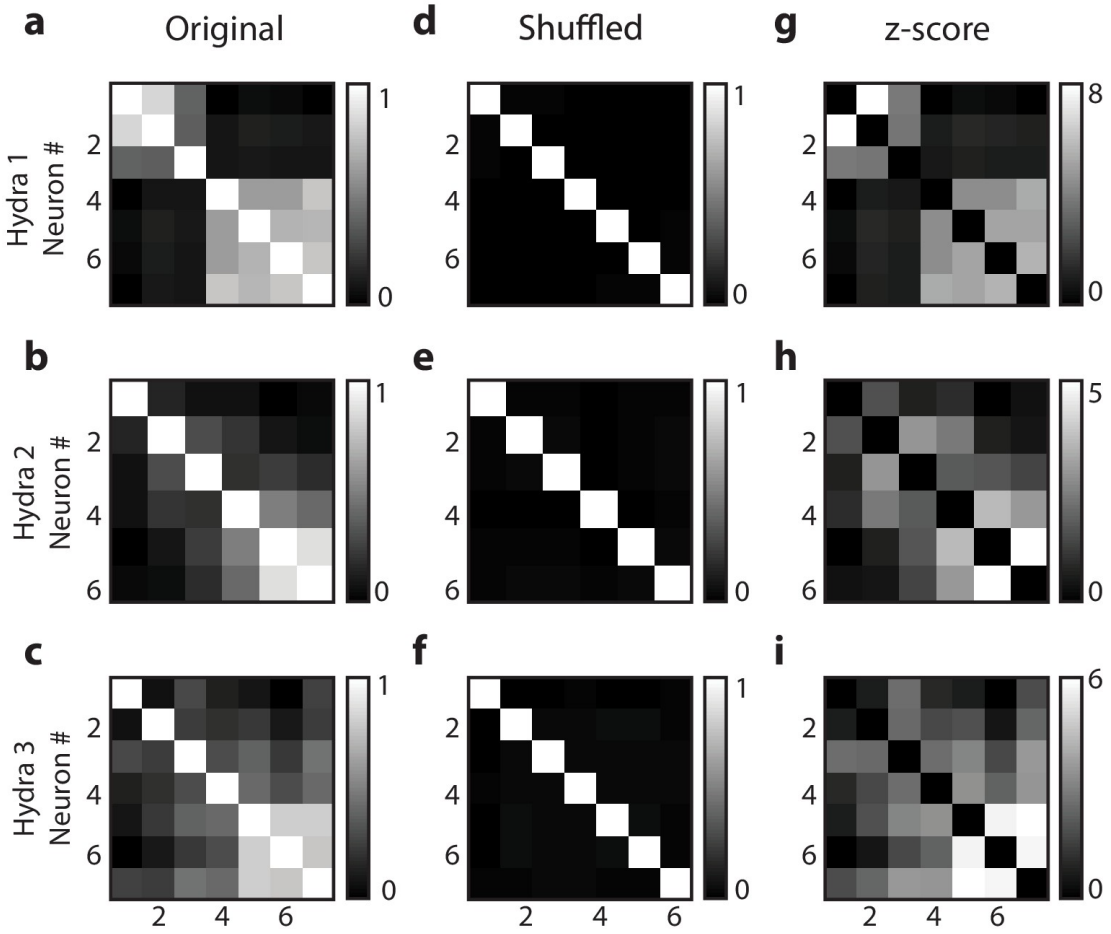


Figure C.21. Random shuffling of single-cell correlation analysis. Individual neuronal fluorescence time series were randomly shuffled to show the correlation coefficients were not due to random chance in all three *Hydra* (each row is a different *Hydra*). Original correlation coefficients from 10 min of raw fluorescence time series (A–C), shuffled correlation coefficients from randomly reshuffled blocks reconstructing 10 min of fluorescence time series (D, E) and z-score of correlation coefficients calculated based on random reshuffling (G–I). Original correlation coefficients were calculated from 10 min of raw fluorescence activity for each of the neurons during mechanical stimulation. Shuffled correlation coefficients were calculated by taking the mean of correlation coefficients from randomly reshuffled fluorescence time series. Briefly, the raw fluorescence time series from 10 min of mechanical stimulation was divided into 100 blocks, which were then randomly recombined for each of the neurons to calculated correlation coefficients and this random reshuffling was repeated 1000 times. Z-score was calculated using the mean and standard deviation of the correlation coefficients calculated from random reshuffling repeated 1000 times. The heat map for z-score (G–I) resembles the original correlation coefficients (A–C), indicating the correlation coefficients of the raw fluorescence time series were not a result of random chance. Except for the diagonals that represent autocorrelation, the mean correlation coefficients resulting from randomly reshuffled fluorescence (E, F) are zero for all neuron pairs, indicating the correlation due to random chance. Description of videos.

Supplemental tables and data can be found at DOI: 10.7554/eLife.64108

CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation describes an in-depth examination of the molecular diversity of the *Hydra vulgaris* nervous system. **Chapter 1** contains a manuscript published in *Science* on which I am a co-author. This study aimed to create a comprehensive single-cell RNA-seq atlas of all cell states in the adult *Hydra* polyp. Although almost all *Hydra* cell types have been morphologically defined, the number of unique transcriptional states and molecular regulation driving cell differentiation were unknown. To address these gaps in knowledge, we sequenced ~25,000 single cell transcriptomes from whole *Hydra* using Drop-seq. Additionally, we generated two neuron-enriched single-cell RNA-seq libraries using FACS to collect cells from a transgenic line expressing GFP in neurons and neuronal progenitors. Using these data, we generated the first molecular map of the *Hydra* nervous system and identified 11 transcriptionally distinct neuronal subtypes. Next, we performed trajectory analysis on *Hydra* interstitial cells and identified a putative progenitor state shared by neurons and gland cells. Finally, we established ATAC-seq in *Hydra* and used the chromatin accessibility data to predict cell fate regulators.

Chapter 2 contains a preprint manuscript on which I am the first author. These results build on the work started in our previously published molecular map of the *Hydra* nervous system (Chapter 1). In this study, we increased the number of sequenced neurons and neural progenitors ten-fold to approximately 35,000 neural single cell transcriptomes, including differentiated neurons and cells undergoing neurogenesis. We surveyed for previously undetected diversity, and in addition to confirming 11 neuron subtypes, we identified the first transcriptional evidence of neurons undergoing transdifferentiation in *Hydra*. We performed ATAC-seq to identify the chromatin state of *Hydra* neurons and identified 48 transcription factors expressed specifically in

the *Hydra* nervous system. This includes transcription factors expressed in all neurons as well as more selectively expressed transcription factors that we hypothesize act as terminal selectors. Finally, we built and validated differentiation trajectories describing the transcriptional changes that underlie *Hydra* neurogenesis. This work represents the most detailed molecular description of the adult *Hydra* nervous system to date.

The **appendix** contains three manuscripts that resulted from collaborative efforts stemming from my thesis work. **Appendix A** is a preprint manuscript on which I am a co-author. This study aimed to answer a fundamental question about neural circuit regeneration—do naïve or terminally differentiated neurons join regenerating neural circuits? We developed a dual-expression transgenic *Hydra* line with a cell-type specific tdTomato in ec5 peduncle neurons and a pan-neuronally expressed GCaMP7s. Using this line, we performed live imaging during peduncle regeneration to capture neural activity while simultaneously tracking the reappearance of ec5 neurons (as determined by tdTomato expression). We found that ec5 neurons terminally differentiate prior to re-establishing their functional role in the nervous system.

Appendix B is a manuscript published in *Genome Research* on which I am a co-author. This study aimed to generate new resources for characterizing transcriptional regulation in *Hydra*, including the following: new genome assemblies for the AEP strain of *Hydra vulgaris*, an updated whole-animal single-cell RNA-seq atlas, and genome-wide maps of chromatin interactions, chromatin accessibility, sequence conservation, and histone modifications. We generated a chromosome-level genome and identified several candidate regulators of cell type-specific transcription.

Appendix C is a manuscript published in *eLife* on which I am coauthor. This study aimed to characterize how *Hydra*'s diffuse nervous system coordinates its mechanosensory response and accomplished this by using microfluidic devices, fluorescent calcium imaging, and surgical resectioning. We found that *Hydra*'s sensorimotor response relies on at least two distinct neural networks in the oral and aboral regions of the animal. Different activity patterns arise in these networks depending on whether the animal is contracting spontaneously or in response to mechanical stimulation.

Together, these papers have established the foundation for the first molecular model of nervous system development in *Hydra*. We have established protocols for generating and isolating transgenic neurons and for using sorted neurons to make enriched, next generation sequencing libraries. We have identified 11 neuron subtypes in *Hydra*, as well as two distinct progenitor populations, each giving rise to populations of endodermal or ectodermal neurons. In addition to the molecular markers that define each neural subtype, we have assigned pseudotime to genes expressed during neuron differentiation, allowing for temporal ordering.

The data presented in this dissertation are hypothesis generators and critical starting points for functional studies of nervous system development, regeneration, and function. For example, we have identified 48 transcription factors expressed specifically in *Hydra*'s neurons. We hypothesize several of these act as terminal selectors, including *neurog1/2/3* (*G008286*), *klf3* (*G022927)*, and *litaf* (*G007969)*, which are expressed in en2, en3, and ec2 neurons, respectively. We can generate and test putative gene regulatory networks by first temporally ordering transcription factors of interest using the differentiation trajectory generated in Chapter 2. We

can then perform transcription factor binding enrichment with the neuron-enriched ATAC-seq data collected in Chapter 2 to narrow down likely candidates. Finally, we can functionally test transcription factors of interest via RNAi and perform downstream validation using molecular markers identified in Chapters 1 and 2.

An outstanding question in regenerative biology is whether regeneration recapitulates development. To answer this, the next step in this project is to build regeneration trajectories to compare to the existing homeostatic trajectories that I built during my dissertation work. Single cell RNA-seq libraries can be collected over a regeneration time course, and after building trajectories of neural regeneration, the temporal expression dynamics of all transcription factors that are expressed during this process can be analyzed to identify similarities and key differences with homeostatic differentiation. These transcription factors can then be functionally tested for their role in driving homeostatic and/or regenerative neurogenesis. These trajectories can also be used to determine the relative contributions of neurogenesis and transdifferentiation during both homeostasis and regeneration. In summary, this dissertation establishes *Hydra* as a valuable developmental and regenerative neurobiology model and provides molecular and computational tools to facilitate future breakthroughs.

REFERENCES

Abràmoff DMD, Magalhães P, Ram S. 2004. Image Processing with ImageJ. *Biophotonics international* **11**:36–42.

Ahrens MB, Li JM, Orger MB, Robson DN, Schier AF, Engert F, Portugues R. 2012. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**:471–477.

Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ. 2013. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods* **10**:413–420.

Aimon S, Katsuki T, Jia T, Grosenick L, Broxton M, Deisseroth K, Sejnowski TJ, Greenspan RJ. 2019. Fast near-whole–brain imaging in adult Drosophila during responses to stimuli and behavior. *PLoS biology* **17**:e2006732.

Albertin CB, Medina-Ruiz S, Mitros T, Schmidbaur H, Sanchez G, Wang ZY, Grimwood J, Rosenthal JJC, Ragsdale CW, Simakov O, Rokhsar DS. 2022. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat Commun* **13**:2427. doi:10.1038/s41467-022-29748-w

Alexopoulos H, Böttger A, Fischer S, Levin A, Wolf A, Fujisawa T, Hayakawa S, Gojobori T, Davies JA, David CN, Bacon JP. 2004. Evolution of gap junctions: the missing link? *Current Biology* **14**:R879–R880. doi:10.1016/j.cub.2004.09.067

Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F. 2016. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**:1241–1243. doi:10.1093/bioinformatics/btv715

Arendt D. 2008. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Reviews Genetics* **9**:868–882.

Artiushin G, Sehgal A. 2017. The Drosophila circuitry of sleep–wake regulation. *Current opinion in neurobiology* **44**:243–250.

Augustin R, Franke A, Khalturin K, Kiko R, Siebert S, Hemmrich G, Bosch TCG. 2006. Dickkopf related genes are components of the positional value gradient in Hydra. *Developmental Biology* **296**:62–70. doi:10.1016/j.ydbio.2006.04.003

Augustin R, Schröder K, Murillo Rincón AP, Fraune S, Anton-Erxleben F, Herbst E-M, Wittlieb J, Schwentner M, Grötzinger J, Wassenaar TM, Bosch TCG. 2017. A secreted antibacterial neuropeptide shapes the microbiome of Hydra. *Nat Commun* **8**:698. doi:10.1038/s41467-017-00625-1

Babonis LS, Martindale MQ. 2017. PaxA, but not PaxC, is required for cnidocyte development in the sea anemone Nematostella vectensis. *EvoDevo* **8**:14. doi:10.1186/s13227-017-0077-7

Badhiwala KN, Gonzales DL, Vercosa DG, Avants BW, Robinson JT. 2018. Microfluidics for electrophysiology, imaging, and behavioral analysis of Hydra. *Lab Chip* **18**:2523–2539. doi:10.1039/c8lc00475g

Badhiwala KN, Primack AS, Juliano CE, Robinson JT. 2021. Multiple neuronal networks coordinate Hydra mechanosensory behavior. *Elife* **10**:e64108.

Bargmann CI, Avery L. 1995. Laser killing of cells in Caenorhabditis elegans. *Methods in cell biology* **48**:225–250.

Bargmann CI, Horvitz HR. 1991. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in C. elegans. *Neuron* **7**:729–742.

Beffagna G. 2019. Zebrafish as a Smart Model to Understand Regeneration After Heart Injury: How Fish Could Help Humans. *Front Cardiovasc Med* **6**. doi:10.3389/fcvm.2019.00107

Bejjani RE, Hammarlund M. 2012. Neural regeneration in Caenorhabditis elegans. *Annual review of genetics* **46**:499–513.
Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. 2020. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**:4267. doi:10.1038/s41467-020-18035-1
Bielen H, Oberleitner S, Marcellini S, Gee L, Lemaire P, Bode H, Rupp R, Technau U. 2008. Divergent functions of two ancient Hydra Brachyury paralogues suggest specific roles for their C-terminal domains in tissue fate induction. *Development (Cambridge, England)* **134**:4187–97. doi:10.1242/dev.010173
Blader P, Fischer N, Gradwohl G, Guillemot F, Strähle U. 1997. The activity of Neurogenin1 is controlled by local cues in the zebrafish embryo. *Development* **124**:4557–4569. doi:10.1242/dev.124.22.4557
Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* **49**:D344–D354. doi:10.1093/nar/gkaa977
Bode HR. 1996. The interstitial cell lineage of hydra: A stem cell system that arose early in evolution. *Journal of cell science* **109 (Pt 6)**:1155–64. doi:10.1242/jcs.109.6.1155
Bode HR, Berking S, David CN, Gierer A, Schaller H, Trenkner E. 1973. Quantitative analysis of cell types during growth and morphogenesis in Hydra. *W Roux' Archiv f Entwicklungsmechanik* **171**:269–285. doi:10.1007/BF00577725
Bode HR, Heimfeld S, Koizumi O, Littlefield C, Yaross M. 1988. Maintenance and regeneration of the nerve net in hydra. *American zoologist* 1053–1063.
Bode HR. 2003. Head regeneration in Hydra. *Developmental dynamics: an official publication of the American Association of Anatomists* **226**:225–236.
Bode HR. 1992. Continuous conversion of neuron phenotype in hydra. *Trends in Genetics* **8**:279–284. doi:10.1016/0168-9525(92)90254-2
Bode HR, Flick KM. 1976. Distribution and dynamics of nematocyte populations in Hydra attenuata. *Journal of Cell Science* **21**:15–34.
Bode HR, Heimfeld S, Chow MA, Huang LW. 1987. Gland cells arise by differentiation from interstitial cells in Hydra attenuata. *Developmental biology* **122**:577–585.
Bode HR, Heimfeld S, Koizumi O, LlTTLEFIELD CL, Yaross MS. 1988. Maintenance and Regeneration of the Nerve Net in Hydra. *Integr Comp Biol* **28**:1053–1063. doi:10.1093/icb/28.4.1053
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170
Bosch TC, Anton-Erxleben F, Hemmrich G, Khalturin K. 2010. The Hydra polyp: nothing but an active stem cell community. *Development, growth & differentiation* **52**:15–25.
Bosch TC, David CN. 1987. Stem cells of Hydra magnipapillata can differentiate into somatic cells and germ line cells. *Developmental biology* **121**:182–191.
Bosch TC, Krylow SM, Bode HR, Steele RE. 1988. Thermotolerance and synthesis of heat shock proteins: these responses are present in Hydra attenuata but absent in Hydra

oligactis. *Proceedings of the National Academy of Sciences* **85**:7927–7931. doi:10.1073/pnas.85.21.7927
Bosch TCG, Klimovich A, Domazet-Lošo T, Gründer S, Holstein TW, Jékely G, Miller DJ, Murillo-Rincon AP, Rentzsch F, Richards GS, Schröder K, Technau U, Yuste R. 2017. Back to the Basics: Cnidarians Start to Fire. *Trends in Neurosciences* **40**:92–105. doi:10.1016/j.tins.2016.11.005
Bradshaw B, Thompson K, Frank U. 2015. Distinct mechanisms underlie oral vs aboral regeneration in the cnidarian Hydractinia echinata. *Elife* **4**:e05506.
Brette R. 2012. Computing with neural synchrony. *PLoS computational biology* **8**:e1002561.
Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**:eaar5780. doi:10.1126/science.aar5780
Broun M, Gee L, Reinhardt B, Bode HR. 2005. Formation of the head organizer in hydra involves the canonical Wnt pathway. *Development* **132**:2907–2916. doi:10.1242/dev.01848
Broussard GJ, Liang R, Tian L. 2014. Monitoring activity in neural circuits with genetically encoded indicators. *Frontiers in molecular neuroscience* **7**:97.
Browne EN. 1909. The production of new hydranths in Hydra by the insertion of small grafts. *Journal of Experimental Zoology* **7**:1–23.
Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**:lqaa108.
Brunet J-P, Tamayo P, Golub TR, Mesirov JP. 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences* **101**:4164–4169. doi:10.1073/pnas.0308531101
Buenrostro J, Wu B, Chang H, Greenleaf W. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**:21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109
Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**:1213–1218.
Burnett AL, Diehl NA. 1964. The nervous system of Hydra. I. Types, distribution and origin of nerve elements. *Journal of Experimental Zoology* **157**:217–226.
Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**:411–420. doi:10.1038/nbt.4096
Bylund M, Andersson E, Novitch BG, Muhr J. 2003. Vertebrate neurogenesis is counteracted by Sox1–3 activity. *Nat Neurosci* **6**:1162–1168. doi:10.1038/nn1131
Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**:e4958. doi:10.7717/peerj.4958
Campbell R, Josephson R, Schwab W, Rushforth N. 1976. Excitability of nerve-free hydra. *Nature* **262**:388–390.
Campbell RD. 1979. Development of hydra lacking interstitial and nerve cells ("epithelial hydra"). *Determinants of spatial organization* 267–293.
Campbell RD. 1976. Elimination of Hydra interstitial and nerve cells by means of colchicine. *Journal of Cell Science* **21**:1–13.

Campbell RD. 1974. Cell Movements in Hydra. *American Zoologist* **14**:523–535.
Campbell Richard D. 1967. Tissue dynamics of steady state growth in Hydra littoralis. II. Patterns of tissue movement. *Journal of Morphology* **121**:19–28.
Campbell Richard D. 1967. Tissue dynamics of steady state growth inHydra littoralis. II. Patterns of tissue movement. *J Morphol* **121**:19–28. doi:10.1002/jmor.1051210103
Cazet J, Juliano C. 2020. Oral Regeneration Is the Default Pathway Triggered by Injury in *Hydra* (preprint). Developmental Biology. doi:10.1101/2020.07.06.189811
Cazet JF, Cho, Adrienne, Juliano C. 2021. Generic injuries are sufficient to induce ectopic Wnt organizers n Hydra. *eLife*.
Cazet JF, Siebert S, Morris Little H, Bertemes P, Primack AS, Ladurner P, Achrainer M, Fredriksen MT, Moreland RT, Singh S, Zhang S, Wolfsberg TG, Schnitzler CE, Baxevanis AD, Simakov O, Hobmayer B, Juliano CE. 2023. A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res* gr.277040.122. doi:10.1101/gr.277040.122
Chalfie M, Sulston JE, White JG, Southgate E, Thomson JN, Brenner S. 1985. The neural circuit for touch sensitivity in Caenorhabditis elegans. *Journal of Neuroscience* **5**:956–964.
Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, Disbennett K, Pfannkoch C, Sumin N, Sutton GG, Viswanathan LD, Walenz B, Goodstein DM, Hellsten U, Kawashima T, Prochnik SE, Putnam NH, Shu S, Blumberg B, Dana CE, Gee L, Kibler DF, Law L, Lindgens D, Martinez DE, Peng J, Wigge PA, Bertulat B, Guder C, Nakamura Y, Ozbek S, Watanabe H, Khalturin K, Hemmrich G, Franke A, Augustin R, Fraune S, Hayakawa E, Hayakawa S, Hirose M, Hwang JS, Ikeo K, Nishimiya-Fujisawa C, Ogura A, Takahashi T, Steinmetz PRH, Zhang X, Aufschnaiter R, Eder M-K, Gorny A-K, Salvenmoser W, Heimberg AM, Wheeler BM, Peterson KJ, Böttger A, Tischler P, Wolf A, Gojobori T, Remington KA, Strausberg RL, Venter JC, Technau U, Hobmayer B, Bosch TCG, Holstein TW, Fujisawa T, Bode HR, David CN, Rokhsar DS, Steele RE. 2010. The dynamic genome of Hydra. *Nature* **464**:592–596. doi:10.1038/nature08830
Chari T, Weissbourd B, Gehring J, Ferraioli A, Leclère L, Herl M, Gao F, Chevalier S, Copley RR, Houliston E. 2021a. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across Clytia medusa cell types. *Science Advances* **7**:eabh1683.
Chari T, Weissbourd B, Gehring J, Ferraioli A, Leclère L, Herl M, Gao F, Chevalier S, Copley RR, Houliston E, Anderson DJ, Pachter L. 2021b. Whole Animal Multiplexed Single-Cell RNA-Seq Reveals Plasticity of *Clytia* Medusa Cell Types (preprint). Systems Biology. doi:10.1101/2021.01.22.427844
Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**:295–300.
Chen X, Mu Y, Hu Y, Kuan AT, Nikitchenko M, Randlett O, Chen AB, Gavornik JP, Sompolinsky H, Engert F. 2018. Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish. *Neuron* **100**:876–890.
Clark DA, Freifeld L, Clandinin TR. 2013. Mapping and cracking sensorimotor circuits in genetic model organisms. *Neuron* **78**:583–595.
Cong L, Wang Z, Chai Y, Hang W, Shang C, Yang W, Bai L, Du J, Wang K, Wen Q. 2017. Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). *Elife* **6**:e28158.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**:959–962. doi:10.1038/nmeth.4396

Dana CE, Glauber KM, Chan TA, Bridge DM, Steele RE. 2012. Incorporation of a horizontally transferred gene into an operon during cnidarian evolution. *PLoS One* **7**:e31643.

Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, Tsegaye G, Tsang A, Wong A, Patel R. 2019. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature methods* **16**:649–657.

Darling JA, Reitzel AR, Burton PM, Mazza ME, Ryan JF, Sullivan JC, Finnerty JR. 2005. Rising starlet: the starlet sea anemone, Nematostella vectensis. *Bioessays* **27**:211–221.

Darmer D, Hauser F, Nothacker H-P, Bosch TCG, Williamson M, Grimmelikhuijzen CJP. 1998. Three different prohormones yield a variety of Hydra-RFamide (Arg-Phe-NH2) neuropeptides in Hydra magnipapillata. *Biochemical Journal* **332**:403–412. doi:10.1042/bj3320403

David CN. 2012. Interstitial stem cells in Hydra: multipotency and decision-making. *Int J Dev Biol* **56**:489–497. doi:10.1387/ijdb.113476cd

David CN, Gierer A. 1974. Cell Cycle Kinetics and Development of Hydra attenuata. III. Nerve and nematocyte differentiation.

David CN, Murphy S. 1977. Characterization of interstitial stem cells in hydra by cloning. *Developmental biology* **58**:372–383.

David CN, Özbek S, Adamczyk P, Meier S, Pauly B, Chapman J, Hwang JS, Gojobori T, Holstein TW. 2008. Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trends in genetics* **24**:431–438.

Davis LE, Burnett AL, Haynes JF, Mumaw VR. 1966. A histological and ultrastructural study of dedifferentiation and redifferentiation of digestive and gland cells in Hydra viridis. *Developmental Biology* **14**:307–329. doi:10.1016/0012-1606(66)90018-2

De Fruyt N, Yu AJ, Rankin CH, Beets I, Chew YL. 2020. The role of neuropeptides in learning: Insights from C. elegans. *The International Journal of Biochemistry & Cell Biology* **125**:105801. doi:10.1016/j.biocel.2020.105801

De Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I. 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proc Natl Acad Sci USA* **110**. doi:10.1073/pnas.1311818110

Della Gaspera B, Weill L, Chanoine C. 2022. Evolution of Somite Compartmentalization: A View From Xenopus. *Frontiers in Cell and Developmental Biology* 3571.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Dohrmann M, Wörheide G. 2017. Dating early animal evolution using phylogenomic data. *Scientific reports* **7**:3599.

Doitsidou M, Minevich G, Kroll JR, Soete G, Gowtham S, Korswagen HC, Zon JS van, Hobert O. 2018. A Caenorhabditis elegans Zinc Finger Transcription Factor, ztf-6, Required for the Specification of a Dopamine Neuron-Producing Lineage. *G3: Genes, Genomes, Genetics* **8**:17–26. doi:10.1534/g3.117.300132

Domazet-Lošo T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in Genetics* **23**:533–539.
Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**:815–818.
Dübel S. 1989. Cell differentiation in the head of Hydra. *Differentiation* **41**:99–109. doi:10.1111/j.1432-0436.1989.tb00737.x
DuBuc TQ, Schnitzler CE, Chrysostomou E, McMahon ET, Febrimarsa, Gahan JM, Buggie T, Gornik SG, Hanley S, Barreira SN, Gonzalez P, Baxevanis AD, Frank U. 2020. Transcription factor AP2 controls cnidarian germ cell induction. *Science* **367**:757–762. doi:10.1126/science.aay6782
Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**:92–95. doi:10.1126/science.aal3327
Dunn CW, Giribet G, Edgecombe GD, Hejnol A. 2014. Animal Phylogeny and Its Evolutionary Implications. *Annual Review of Ecology, Evolution, and Systematics* **45**:371–395. doi:10.1146/annurev-ecolsys-120213-091627
Dunn CW, Howison M, Zapata F. 2013. Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* **14**:330. doi:10.1186/1471-2105-14-330
Dunne JF, Javois LC, Huang LW, Bode HR. 1985. A subset of cells in the nerve net of Hydra oligactis defined by a monoclonal antibody: Its arrangement and development. *Developmental Biology* **109**:41–53. doi:10.1016/0012-1606(85)90344-6
Dupre C, Yuste R. 2017a. Non-overlapping Neural Networks in Hydra vulgaris. *Current Biology* **27**:1085–1097. doi:10.1016/j.cub.2017.02.049
Dupre C, Yuste R. 2017b. Non-overlapping Neural Networks in Hydra vulgaris. *Current Biology* **27**:1085–1097. doi:10.1016/j.cub.2017.02.049
Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *cels* **3**:95–98. doi:10.1016/j.cels.2016.07.002
Dworkin S, Heath JK, deJong-Curtain TA, Hogan BM, Lieschke GJ, Malaterre J, Ramsay RG, Mantamadiotis T. 2007. CREB activity modulates neural cell proliferation, midbrain–hindbrain organization and patterning in zebrafish. *Developmental Biology* **307**:127–141. doi:10.1016/j.ydbio.2007.04.026
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**:1–14.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**:1–14.
Endl I, Lohmann JU, Bosch TCG. 1999. Head-specific gene expression in Hydra: Complexity of DNA– protein interactions at the promoter of ks1 is inversely correlated to the head activation potential. *Proceedings of the National Academy of Sciences* **96**:1445–1450. doi:10.1073/pnas.96.4.1445
Engel U, Oezbek S, Engel R, Petri B, Lottspeich F, Holstein TW. 2002. Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in *Hydra*. *Journal of Cell Science* **115**:3923–3934. doi:10.1242/jcs.00084

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE* **7**:e47768. doi:10.1371/journal.pone.0047768

Epp L, Tardent P. 1978. The distribution of nerve cells in Hydra attenuata Pall. *Wilhelm Roux's archives of developmental biology* **185**:185–193.

Erwin DH. 2009. Early origin of the bilaterian developmental toolkit. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**:2253–2261. doi:10.1098/rstb.2009.0038

Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**:eaar3131. doi:10.1126/science.aar3131

Fedders H, Augustin R, Bosch TCG. 2004. A Dickkopf-3-related gene is expressed in differentiating nematocytes in the basal metazoan Hydra. *Dev Genes Evol* **214**:72–80. doi:10.1007/s00427-003-0378-9

Ferenc J, Papasaikas P, Ferralli J, Nakamura Y, Smallwood S, Tsiairis CD. 2021. Mechanical oscillations orchestrate axial patterning through Wnt activation in Hydra. *Science Advances* **7**:eabj6897.

Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. 2018. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* **360**:eaaq1736. doi:10.1126/science.aaq1736

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**:D279–D285. doi:10.1093/nar/gkv1344

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**:9451–9457. doi:10.1073/pnas.1921046117

Fornes O, Castro-Mondragon JA, Khan A, Van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **48**:D87–D92.

Fourrage C, Swann K, Gonzalez Garcia JR, Campbell AK, Houliston E. 2014. An endogenous green fluorescent protein–photoprotein pair in Clytia hemisphaerica eggs shows co-targeting to mitochondria and efficient bioluminescence energy transfer. *Open Biology* **4**:130206.

Fraune J, Alsheimer M, Volff J-N, Busch K, Fraune S, Bosch TCG, Benavente R. 2012. Hydra meiosis reveals unexpected conservation of structural synaptonemal complex proteins across metazoans. *Proceedings of the National Academy of Sciences* **109**:16588–16593. doi:10.1073/pnas.1206875109

Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB, Samoilovich MP, Bosch TCG. 2010. In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proceedings of the National Academy of Sciences* **107**:18067–18072. doi:10.1073/pnas.1008573107

Fraune S, Bosch TCG. 2007. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. *Proceedings of the National Academy of Sciences* **104**:13146–13151. doi:10.1073/pnas.0703375104

Freitas R, Zhang G, Cohn MJ. 2006. Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature* **442**:1033–1037.

Fruchterman TMJ, Reingold EM. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience* **21**:1129–1164. doi:10.1002/spe.4380211102

Fujisawa T, David CN. 1981. Commitment during nematocyte differentiation in Hydra. *Journal of Cell Science* **48**:207–222.

Galliot B, Quiquand M, Ghila L, de Rosa R, Miljkovic-Licina M, Chera S. 2009. Origins of neurogenesis, a cnidarian view. *Developmental Biology*, Special Section: Evolution of Developmental Regulatory Systems **332**:2–24. doi:10.1016/j.ydbio.2009.05.563

Gan L, Xiang M, Zhou L, Wagner DS, Klein WH, Nathans J. 1996. POU domain factor Brn-3b is required for the development of a large set of retinal ganglion cells. *Proceedings of the National Academy of Sciences* **93**:3920–3925. doi:10.1073/pnas.93.9.3920

Gandhi AV, Mosser EA, Oikonomou G, Prober DA. 2015. Melatonin is required for the circadian regulation of sleep. *Neuron* **85**:1193–1199.

Gee L, Hartig J, Law L, Wittlieb J, Khalturin K, Bosch TC, Bode HR. 2010. β-catenin plays a central role in setting up the head organizer in hydra. *Developmental biology* **340**:116–124.

Ghosh DD, Nitabach MN, Zhang Y, Harris G. 2017. Multisensory integration in C. elegans. *Current opinion in neurobiology* **43**:110–118.

Gierer A, Berking S, Bode H, David CN, Flick K, Hansmann G, Schaller H, Trenkner E. 1972. Regeneration of hydra from reaggregated cells. *Nat New Biol* **239**:98–101. doi:10.1038/newbio239098a0

Gierer A, Meinhardt H. 1972. A theory of biological pattern formation. *Kybernetik* **12**:30–39.

Gilbert D. 2013. Gene-omes built from mRNA seq not genome DNA.

Glauber KM, Dana CE, Park SS, Colby DA, Noro Y, Fujisawa T, Chamberlin AR, Steele RE. 2013. A small molecule screen identifies a novel compound that induces a homeotic transformation in Hydra. *Development* **140**:4788–4796. doi:10.1242/dev.094490

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* **108**:1513–1518. doi:10.1073/pnas.1017351108

Goel T, Wang R, Martin S, Lanphear E, Collins E-MS. 2019. Linalool acts as a fast and reversible anesthetic in Hydra. *PloS one* **14**:e0224221.

Gold DA, Jacobs DK. 2013. Stem cell dynamics in Cnidaria: are there unifying principles? *Development genes and evolution* **223**:53–66.

Gonzales D, Zhou J, Fan B, Robinson J. 2019. A microfluidic-induced C. elegans sleep state. Nat. Commun. 10: 5035.

Gonzales DL, Badhiwala KN, Avants BW, Robinson JT. 2020. Bioelectronics for Millimeter-Sized model organisms. *Iscience* **23**:100917.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-

length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**:644–652. doi:10.1038/nbt.1883
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**:1017–1018.
Grapin-Botton A, Constam D. 2007. Evolution of the mechanisms and molecular control of endoderm formation. *Mechanisms of Development* **124**:253–278. doi:10.1016/j.mod.2007.01.001
Greber MJ, David CN, Holstein TW. 1992. A quantitative method for separation of living Hydra cells. *Roux's Arch Dev Biol* **201**:296–300. doi:10.1007/BF00592110
Grens A, Gee L, Fisher DA, Bode HR. 1996. CnNK-2,an NK-2 Homeobox Gene, Has a Role in Patterning the Basal End of the Axis in Hydra. *Developmental Biology* **180**:473–488. doi:10.1006/dbio.1996.0321
Guder C, Pinho S, Nacak TG, Schmidt HA, Hobmayer B, Niehrs C, Holstein TW. 2006. An ancient Wnt-Dickkopf antagonism in Hydra. *Development* **133**:901–911. doi:10.1242/dev.02265
Gufler S, Artes B, Bielen H, Krainer I, Eder M-K, Falschlunger J, Bollmann A, Ostermann T, Valovka T, Hartl M. 2018. β-Catenin acts in a position-independent regeneration response in the simple eumetazoan Hydra. *Developmental biology* **433**:310–323.
Gumucio DL, Heilstedt-Williamson H, Gray TA, Tarlé SA, Shelton DA, Tagle DA, Slightom JL, Goodman M, Collins FS. 1992. Phylogenetic Footprinting Reveals a Nuclear Protein Which Binds to Silencer Sequences in the Human y and e Globin Genes. *Molecular and Cellular Biology* **12**:4919–4929. doi:10.1128/mcb.12.11.4919-4929.1992
Guo F, Holla M, Díaz MM, Rosbash M. 2018. A circadian output circuit controls sleep-wake arousal in Drosophila. *Neuron* **100**:624–635.
Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**:5654–5666. doi:10.1093/nar/gkg770
Haesemeyer M, Robson DN, Li JM, Schier AF, Engert F. 2018. A brain-wide circuit model of heat-evoked swimming behavior in larval zebrafish. *Neuron* **98**:817–831.
Hafemeister C, Satija R. 2019a. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome biology* **20**:296.
Hafemeister C, Satija R. 2019b. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**:296. doi:10.1186/s13059-019-1874-1
Hager G, David CN. 1997. Pattern of differentiated nerve cells in hydra is determined by precursor migration. *Development* **124**:569–576.
Haghverdi L, Buettner F, Theis FJ. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**:2989–2998. doi:10.1093/bioinformatics/btv325
Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**:845–848. doi:10.1038/nmeth.3971
Hahne F, Ivanek R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor In: Mathé E, Davis S, editors. Statistical Genomics: Methods and Protocols, Methods in Molecular Biology. New York, NY: Springer. pp. 335–351. doi:10.1007/978-1-4939-3578-9_16

Hamada M, Satoh N, Khalturin K. 2020. A Reference Genome from the Symbiotic Hydrozoan, Hydra viridissima. *G3 Genes|Genomes|Genetics* **10**:3883–3895. doi:10.1534/g3.120.401411

Hamada M, Schröder K, Bathia J, Kürn U, Fraune S, Khalturina M, Khalturin K, Shinzato C, Satoh N, Bosch TC. 2018. Metabolic co-dependence drives the evolutionarily ancient Hydra–Chlorella symbiosis. *Elife* **7**:e35122.

Hansen GN, Williamson M, Grimmelikhuijzen CJ. 2000. Two-color double-labeling in situ hybridization of whole-mount Hydra using RNA probes for five different Hydra neuropeptide preprohormones: evidence for colocalization. *Cell and tissue research* **301**:245–253.

Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* **38**:e131. doi:10.1093/nar/gkq224

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M. 2021a. Integrated analysis of multimodal single-cell data. *Cell* **184**:3573-3587. e29.

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. 2021b. Integrated analysis of multimodal single-cell data. *Cell* **184**:3573-3587.e29. doi:10.1016/j.cell.2021.04.048

Harland RM. 2018. A new view of embryo development and regeneration. *Science* **360**:967–968. doi:10.1126/science.aat8413

Hartigan JA, Hartigan PM. 1985. The Dip Test of Unimodality. *The Annals of Statistics* **13**:70–84.

Hartl M, Glasauer S, Gufler S, Raffeiner A, Puglisi K, Breuker K, Bister K, Hobmayer B. 2019. Differential regulation of myc homologs by Wnt/β-Catenin signaling in the early metazoan Hydra. *The FEBS Journal* **286**:2295–2310.

Hayakawa E, Fujisawa C, Fujisawa T. 2004. Involvement of Hydra achaete–scute gene CnASH in the differentiation pathway of sensory neurons in the tentacles. *Dev Genes Evol* **214**:486–492. doi:10.1007/s00427-004-0430-4

Hecker A, Schulze W, Oster J, Richter DO, Schuster S. 2020. Removing a single neuron in a vertebrate brain forever abolishes an essential behavior. *Proceedings of the National Academy of Sciences* **117**:3254–3260. doi:10.1073/pnas.1918578117

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**:576–589. doi:10.1016/j.molcel.2010.05.004

Hemmrich G, Khalturin K, Boehm A-M, Puchert M, Anton-Erxleben F, Wittlieb J, Klostermeier UC, Rosenstiel P, Oberg H-H, Domazet-Lošo T, Sugimoto T, Niwa H, Bosch TCG. 2012. Molecular Signatures of the Three Stem Cell Lineages in Hydra and the Emergence of Stem Cell Function at the Base of Multicellularity. *Molecular Biology and Evolution* **29**:3267–3280. doi:10.1093/molbev/mss134

Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**:1341–1342. doi:10.1093/bioinformatics/btt128

Hill AJ, Mansfield R, Lopez JMNG, Raizen DM, Van Buskirk C. 2014. Cellular Stress Induces a Protective Sleep-like State in C. elegans. *Current Biology* **24**:2399–2405. doi:10.1016/j.cub.2014.08.040

Hobert O. 2016. Chapter Twenty-Five - Terminal Selectors of Neuronal Identity In: Wassarman PM, editor. Current Topics in Developmental Biology, Essays on Developmental Biology, Part A. Academic Press. pp. 455–475. doi:10.1016/bs.ctdb.2015.12.007

Hobmayer B, Jenewein M, Eder D, Eder M-K, Glasauer S, Gufler S, Hartl M, Salvenmoser W. 2012. Stemness in Hydra - a current perspective. *Int J Dev Biol* **56**:509–517. doi:10.1387/ijdb.113426bh

Hobmayer B, Rentzsch F, Kuhn K, Happel CM, von Laue CC, Snyder P, Rothbächer U, Holstein TW. 2000. WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. *Nature* **407**:186–189. doi:10.1038/35025063

Hobmayer E, David C. 1989. Differentiation of a Nerve Cell-Battery Cell Complex in HydraEvolution of the First Nervous Systems. Springer US, Boston, MA. pp. 71–80.

Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen RGHP, van den Broek B, Mhaskar AN, Teunissen H, St Hilaire BG, Weisz D, Omer AD, Pham M, Colaric Z, Yang Z, Rao SSP, Mitra N, Lui C, Yao W, Khan R, Moroz LL, Kohn A, St. Leger J, Mena A, Holcroft K, Gambetta MC, Lim F, Farley E, Stein N, Haddad A, Chauss D, Mutlu AS, Wang MC, Young ND, Hildebrandt E, Cheng HH, Knight CJ, Burnham TLU, Hovel KA, Beel AJ, Mattei P-J, Kornberg RD, Warren WC, Cary G, Gómez-Skarmeta JL, Hinman V, Lindblad-Toh K, Di Palma F, Maeshima K, Multani AS, Pathak S, Nel-Themaat L, Behringer RR, Kaur P, Medema RH, van Steensel B, de Wit E, Onuchic JN, Di Pierro M, Lieberman Aiden E, Rowland BD. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**:984–989. doi:10.1126/science.abe2218

Hoffmeister S. 1991. Analysis of a foot regeneration deficient strain of Hydra oligactis. *Mechanisms of development* **35**:181–192.

Holstein T. 1981. The morphogenesis of nematocytes in Hydra and Forsklia: An ultrastructural study. *Journal of Ultrastructure Research* **75**:276–290. doi:10.1016/S0022-5320(81)80085-8

Holstein TW, Hobmayer E, David CN. 1991. Pattern of epithelial cell cycling in hydra. *Developmental Biology* **148**:602–611. doi:10.1016/0012-1606(91)90277-A

Huang L, Ledochowitsch P, Knoblich U, Lecoq J, Murphy GJ, Reid RC, De Vries SE, Koch C, Zeng H, Buice MA, Waters J, Li L. 2021. Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife* **10**:e51675. doi:10.7554/eLife.51675

Hufnagel LA, Kass-Simon G. 2016. The two nerve rings of the hypostomal nervous system of Hydra vulgaris—an immunohistochemical analysis. *Cell Tissue Res* **366**:255–269. doi:10.1007/s00441-016-2447-1

Hufnagel LA, Kass-Simon G, Lyon MK. 1985. Functional organization of battery cell complexes in tentacles of Hydra attenuata. *Journal of Morphology* **184**:323–341. doi:10.1002/jmor.1051840307

Hulett RE, Kimura JO, Bolaños DM, Luo Y-J, Ricci L, Srivastava M. 2022. Acoel single-cell atlas reveals expression dynamics and heterogeneity of a pluripotent stem cell population (preprint). Developmental Biology. doi:10.1101/2022.02.10.479464

Hulett RE, Potter D, Srivastava M. 2020. Neural architecture and regeneration in the acoel Hofstenia miamia. *Proceedings of the Royal Society B: Biological Sciences* **287**:20201198. doi:10.1098/rspb.2020.1198

Hwang JS, Ohyanagi H, Hayakawa S, Osato N, Nishimiya-Fujisawa C, Ikeo K, David CN, Fujisawa T, Gojobori T. 2007. The evolutionary emergence of cell type-specific genes inferred from the gene expression analysis of Hydra. *Proceedings of the National Academy of Sciences* **104**:14735–14740. doi:10.1073/pnas.0703331104

Hwang JS, Takaku Y, Chapman J, Ikeo K, David CN, Gojobori T. 2008. Cilium Evolution: Identification of a Novel Protein, Nematocilin, in the Mechanosensory Cilium of Hydra Nematocytes. *Molecular Biology and Evolution* **25**:2009–2017. doi:10.1093/molbev/msn154

Hwang JS, Takaku Y, Momose T, Adamczyk P, Özbek S, Ikeo K, Khalturin K, Hemmrich G, Bosch TCG, Holstein TW, David CN, Gojobori T. 2010. Nematogalectin, a nematocyst protein with GlyXY and galectin domains, demonstrates nematocyte-specific alternative splicing in Hydra. *Proceedings of the National Academy of Sciences* **107**:18539–18544. doi:10.1073/pnas.1003256107

Iachetta R, Ambrosone A, Klimovich A, Wittlieb J, Onorato G, Candeo A, D'andrea C, Intartaglia D, Scotti N, Blasio M. 2018. Real time dynamics of β-catenin expression during Hydra development, regeneration and Wnt signalling activation. *International Journal of Developmental Biology* **62**:311–318.

Inoue C, Bae S, Takatsuka K, Inoue T, Bessho Y, Kageyama R. 2001. Math6, a bHLH gene expressed in the developing nervous system, regulates neuronal versus glial differentiation. *Genes to Cells* **6**:977–986.

Itayama T, Sawada Y. 1995. Development of electrical activity in regenerating aggregates of hydra cells. *J Exp Zool* **273**:519–526. doi:10.1002/jez.1402730608

Jackman SD, Coombe L, Chu J, Warren RL, Vandervalk BP, Yeo S, Xue Z, Mohamadi H, Bohlmann J, Jones SJM, Birol I. 2018. Tigmint: correcting assembly errors using linked reads from large molecules. *BMC Bioinformatics* **19**:393. doi:10.1186/s12859-018-2425-6

Jacomy M, Venturini T, Heymann S, Bastian M. 2014. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE* **9**:e98679. doi:10.1371/journal.pone.0098679

Jarman AP, Grau Y, Jan LY, Jan YN. 1993. atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system. *Cell* **73**:1307–1321. doi:10.1016/0092-8674(93)90358-W

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240. doi:10.1093/bioinformatics/btu031

Juliano CE, Lin H, Steele RE. 2014. Generation of Transgenic Hydra by Embryo Microinjection. *JoVE (Journal of Visualized Experiments)* e51888. doi:10.3791/51888

Juliano CE, Reich A, Liu N, Götzfried J, Zhong M, Uman S, Reenan RA, Wessel GM, Steele RE, Lin H. 2014c. PIWI proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. *Proc Natl Acad Sci USA* **111**:337–342. doi:10.1073/pnas.1320965111

Juliano CE, Swartz SZ, Wessel GM. 2010. A conserved germline multipotency program. *Development* **137**:4113–4126. doi:10.1242/dev.047969

Julier Z, Park AJ, Briquez PS, Martino MM. 2017. Promoting tissue regeneration by modulating the immune system. *Acta Biomaterialia* **53**:13–28. doi:10.1016/j.actbio.2017.01.056
Kakebeen AD, Chitsazan AD, Williams MC, Saunders LM, Wills AE. 2020. Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of regeneration in neural progenitors. *Elife* **9**:e52648.
Kanaya HJ, Park S, Kim J, Kusumi J, Krenenou S, Sawatari E, Sato A, Lee J, Bang H, Kobayakawa Y, Lim C, Itoh TQ. 2020. A sleep-like state in *Hydra* unravels conserved sleep mechanisms during the evolutionary development of the central nervous system. *Sci Adv* **6**:eabb9415. doi:10.1126/sciadv.abb9415
Kaplan HS, Nichols ALA, Zimmer M. 2018. Sensorimotor integration in *Caenorhabditis elegans* : a reappraisal towards dynamic and distributed computations. *Phil Trans R Soc B* **373**:20170371. doi:10.1098/rstb.2017.0371
Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP. 2017. The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**:194–199. doi:10.1126/science.aan3235
Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature communications* **10**:1930.
Kayser MS, Biron D. 2016. Sleep and Development in Genetically Tractable Model Organisms. *Genetics* **203**:21–33. doi:10.1534/genetics.116.189589
Keramidioti A, Schneid S, Busse C, Laue CC von, Bertulat B, Salvenmoser W, Hess M, Alexandrova O, Glauber K, Steele R, Hobmayer B, Holstein TW, David CN. 2023. A new look at the architecture and dynamics of the Hydra nerve net. doi:10.1101/2023.02.22.529525
Khalturin K, Anton-Erxleben F, Sassmann S, Wittlieb J, Hemmrich G, Bosch TCG. 2008. A novel gene family controls species-specific morphological traits in Hydra. *PLoS biology* **6**:e278.
Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TC. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in Genetics* **25**:404–413.
Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46**:D260–D266. doi:10.1093/nar/gkx1126
Khan A, Mathelier A. 2018. JASPAR RESTful API: accessing JASPAR data from any programming language. *Bioinformatics* **34**:1612–1614. doi:10.1093/bioinformatics/btx804
Kim DH, Kim J, Marques JC, Grama A, Hildebrand DGC, Gu W, Li JM, Robson DN. 2017. Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nat Methods* **14**:1107–1114. doi:10.1038/nmeth.4429
Klimovich A, Giacomello S, Björklund Å, Faure L, Kaucka M, Giez C, Murillo-Rincon AP, Matt A-S, Willoweit-Ohl D, Crupi G, De Anda J, Wong GCL, D'Amato M, Adameyko I, Bosch TCG. 2020. Prototypical pacemaker neurons interact with the resident microbiota. *Proc Natl Acad Sci USA* **117**:17854–17863. doi:10.1073/pnas.1920469117
Klimovich A, Wittlieb J, Bosch TC. 2019. Transgenesis in Hydra to characterize gene function and visualize cell behavior. *Nature protocols* **14**:2069–2090.

Knoblich U, Huang L, Zeng H, Li L. 2019. Neuronal cell-subtype specificity of neural synchronization in mouse primary visual cortex. *Nature communications* **10**:2533.
Knoepfler PS, Cheng PF, Eisenman RN. 2002. N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. *Genes Dev* **16**:2699–2712. doi:10.1101/gad.1021202
Koizumi O, Bode HR. 1991. Plasticity in the nervous system of adult hydra. III. Conversion of neurons to expression of a vasopressin-like immunoreactivity depends on axial location. *J Neurosci* **11**:2011–2020. doi:10.1523/JNEUROSCI.11-07-02011.1991
Koizumi O, Bode HR. 1986. Plasticity in the nervous system of adult hydra: I. The position-dependent expression of FMRFamide-like immunoreactivity. *Developmental Biology* **116**:407–421. doi:10.1016/0012-1606(86)90142-9
Koizumi O, Itazawa M, Mizumoto H, Minobe S, Javois LC, Grimmelikhuijzen CJP, Bode HR. 1992. Nerve ring of the hypostome in Hydra. I. Its structure, development, and maintenance. *Journal of Comparative Neurology* **326**:7–21. doi:10.1002/cne.903260103
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology* **37**:540–546.
Kondili M, Fust A, Preussner J, Kuenne C, Braun T, Looso M. 2017a. UROPA: a tool for Universal RObust Peak Annotation. *Sci Rep* **7**:2593. doi:10.1038/s41598-017-02464-y
Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**:722–736. doi:10.1101/gr.215087.116
Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene set enrichment analysis. doi:10.1101/060012
Kortschak RD, Samuel G, Saint R, Miller DJ. 2003. EST Analysis of the Cnidarian Acropora millepora Reveals Extensive Gene Loss and Rapid Sequence Divergence in the Model Invertebrates. *Current Biology* **13**:2190–2195. doi:10.1016/j.cub.2003.11.030
Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA, Sabeti PC. 2019. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**:e43803.
Kroehne V, Freudenreich D, Hans S, Kaslin J, Brand M. 2011. Regeneration of the adult zebrafish brain from neurogenic radial glia-type progenitors. *Development* **138**:4831–4841. doi:10.1242/dev.072587
Krug K, Salzman CD, Waddell S. 2015. Understanding the brain by controlling neural activity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**:20140201. doi:10.1098/rstb.2014.0201
Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**:1813–1831. doi:10.1101/gr.136184.111
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359. doi:10.1038/nmeth.1923

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. doi:10.1186/gb-2009-10-3-r25

Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Péron S, Jager M, Condamine T, Pottin K, Romano S. 2019. The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. *Nature Ecology & Evolution* **3**:801–810.

Lee JE. 1997. Basic helix-loop-helix genes in neural development. *Current Opinion in Neurobiology* **7**:13–20.

Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N, Weintraub H. 1995. Conversion of *Xenopus* Ectoderm into Neurons by NeuroD, a Basic Helix-Loop-Helix Protein. *Science* **268**:836–844. doi:10.1126/science.7754368

Lee-Liu D, Méndez-Olivos EE, Muñoz R, Larraín J. 2017. The African clawed frog Xenopus laevis: a model organism to study regeneration of the central nervous system. *Neuroscience letters* **652**:82–93.

Lemon WC, Pulver SR, Höckendorf B, McDole K, Branson K, Freeman J, Keller PJ. 2015. Whole-central nervous system functional imaging in larval Drosophila. *Nat Commun* **6**:7924. doi:10.1038/ncomms8924

Lengfeld T, Watanabe H, Simakov O, Lindgens D, Gee L, Law L, Schmidt HA, Özbek S, Bode H, Holstein TW. 2009. Multiple Wnts are involved in Hydra organizer formation and regeneration. *Developmental Biology* **330**:186–199. doi:10.1016/j.ydbio.2009.02.004

Lenhoff HM, Brown RD. 1970. Mass culture of hydra: an improved method and its application to other aquatic invertebrates. *Laboratory Animals* **4**:139–154.

Lentz TL, Barrnett RJ. 1965. Fine Structure of the Nervous System of hydra. *American Zoologist* **5**:341–356.

Lesh-Laurie GE. 1983. Separating Viable Tissue Layers In: Lenhoff HM, editor. Hydra: Research Methods. Boston, MA: Springer US. pp. 267–271. doi:10.1007/978-1-4757-0596-6_35

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. doi:10.1186/1471-2105-12-323

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100. doi:10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* **5**. doi:10.1214/11-AOAS466

Li Yunfeng, Gao L, Pan Y, Tian M, Li Yulong, He C, Dong Y, Sun Y, Zhou Z. 2020. Chromosome-level reference genome of the jellyfish Rhopilema esculentum. *GigaScience* **9**:giaa036.

Liu Y, Kossack ME, McFaul ME, Christensen LN, Siebert S, Wyatt SR, Kamei CN, Horst S, Arroyo N, Drummond IA. 2022. Single-cell transcriptome reveals insights into the development and function of the zebrafish ovary. *Elife* **11**:e76014.

Livak KJ, Schmittgen TD. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**:402–408. doi:10.1006/meth.2001.1262

Lobo D, Beane WS, Levin M. 2012. Modeling Planarian Regeneration: A Primer for Reverse-Engineering the Worm. *PLOS Computational Biology* **8**:e1002481. doi:10.1371/journal.pcbi.1002481

Lohman BK, Weber JN, Bolnick DI. 2016. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources* **16**:1315–1321. doi:10.1111/1755-0998.12529

Lohmann JU, Endl I, Bosch TCG. 1999. Silencing of Developmental Genes in Hydra. *Developmental Biology* **214**:211–214. doi:10.1006/dbio.1999.9407

Lommel M, Strompen J, Hellewell AL, Balasubramanian GP, Christofidou ED, Thomson AR, Boyle AL, Woolfson DN, Puglisi K, Hartl M, Holstein TW, Adams JC, Özbek S. 2018. Hydra Mesoglea Proteome Identifies Thrombospondin as a Conserved Component Active in Head Organizer Restriction. *Sci Rep* **8**:11753. doi:10.1038/s41598-018-30035-2

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2020. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**:422–423. doi:10.1093/bioinformatics/btaa692

Lu B, Wang KH, Nose A. 2009. Molecular mechanisms underlying neural circuit formation. *Current Opinion in Neurobiology*, Development **19**:162–167. doi:10.1016/j.conb.2009.04.004

Lu Z, Elliott MR, Chen Y, Walsh JT, Klibanov AL, Ravichandran KS, Kipnis J. 2011. Phagocytic activity of neuronal progenitors regulates adult neurogenesis. *Nat Cell Biol* **13**:1076–1083. doi:10.1038/ncb2299

Lust K, Maynard A, Gomes T, Fleck JS, Camp JG, Tanaka EM, Treutlein B. 2022. Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. *Science* **377**:eabp9262.

Lyon MK, Kass-Simon G, Hufnagel LA. 1982. Ultrastructural analysis of nematocyte removal in Hydra. *Tissue and Cell* **14**:415–424. doi:10.1016/0040-8166(82)90036-2

Macosko Evan Z., Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**:1202–1214. doi:10.1016/j.cell.2015.05.002

Marioni JC, Arendt D. 2017. How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology. *Annual Review of Cell and Developmental Biology* **33**:537–553. doi:10.1146/annurev-cellbio-100616-060818

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**:10.

Martin VJ, Littlefield CL, Archer WE, Bode HR. 1997. Embryogenesis in hydra. *The Biological Bulletin* **192**:345–363.

Martínez D, Iñiguez A, Percell K, Willner J, Signorovitch J, Campbell R. 2010. Phylogeny and biogeography of Hydra (Cnidaria: Hydridae) using mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution* **57**:403–410.

Martinez DE, Dirksen M-L, Bode PM, Jamrich M, Steele RE, Bode HR. 1997. Budhead,a Fork Head/HNF-3 Homologue, Is Expressed during Axis Formation and Head Specification in Hydra. *Developmental Biology* **192**:523–536. doi:10.1006/dbio.1997.8715

Martínez DE. 1998. Mortality patterns suggest lack of senescence in hydra. *Experimental gerontology* **33**:217–225.

Mast SO. 1903. Reactions to temperature changes in Spirillum, Hydra, and fresh-water planarians. *American Journal of Physiology-Legacy Content* **10**:165–190. doi:10.1152/ajplegacy.1903.10.4.165
McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**:861. doi:10.21105/joss.00861
McLeay RC, Bailey TL. 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**:165. doi:10.1186/1471-2105-11-165
McNeil PL. 1981. Mechanisms of nutritive endocytosis. *Journal of Cell Science* **49**:311–339.
Meers MP, Tenenbaum D, Henikoff S. 2019. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & chromatin* **12**:1–11.
Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome biology* **14**:1–20.
Miljkovic-Licina M, Chera S, Ghila L, Galliot B. 2007. Head regeneration in wild-type hydra requires de novo neurogenesis. *Development (Cambridge, England)* **134**:1191–201. doi:10.1242/dev.02804
Miljkovic-Licina M, Gauchat D, Galliot B. 2004. Neuronal evolution: analysis of regulatory genes in a first-evolved nervous system, the hydra nervous system. *Biosystems*, Papers presented at the Fifth International Workshop on Information Processing in Cells and Tissues **76**:75–87. doi:10.1016/j.biosystems.2004.05.030
Mitgutsch C, Hauser F, Grimmelikhuijzen CJP. 1999. Expression and Developmental Regulation of the Hydra-RFamide and Hydra-LWamide Preprohormone Genes inHydra:Evidence for Transient Phases of Head Formation. *Developmental Biology* **207**:189–203. doi:10.1006/dbio.1998.9150
Miyata T, Maeda T, Lee JE. 1999. NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. *Genes Dev* **13**:1647–1652.
Münder S, Käsbauer T, Prexl A, Aufschnaiter R, Zhang X, Towb P, Böttger A. 2010. Notch signalling defines critical boundary during budding in Hydra. *Developmental biology* **344**:331–345.
Murad R, Macias-Muñoz A, Wong A, Ma X, Mortazavi A. 2021. Coordinated gene expression and chromatin regulation during hydra head regeneration. *Genome Biology and Evolution* **13**:evab221.
Muscatine L. 1961. Biology of Hydra. University of Miami Press.
Musser JM, Schippers KJ, Nickel M, Mizzon G, Kohn AB, Pape C, Ronchi P, Papadopoulos N, Tarashansky AJ, Hammel JU. 2021. Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science* **374**:717–723.
Nakanishi Y, Shiratsuchi A. 2004. Phagocytic Removal of Apoptotic Spermatogenic Cells by Sertoli Cells: Mechanisms and Consequences. *Biological & Pharmaceutical Bulletin* **27**:13–16. doi:10.1248/bpb.27.13
Nguyen JP, Shipley FB, Linder AN, Plummer GS, Liu M, Setru SU, Shaevitz JW, Leifer AM. 2016. Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **113**. doi:10.1073/pnas.1507110112
Nichols ALA, Eichler T, Latham R, Zimmer M. 2017. A global brain state underlies *C. elegans* sleep behavior. *Science* **356**:eaam6851. doi:10.1126/science.aam6851
Nishimiya-Fujisawa C, Kobayashi S. 2012. Germline stem cells and sex determination in Hydra. *Int J Dev Biol* **56**:499–508. doi:10.1387/ijdb.123509cf

Nong W, Cao J, Li Y, Qu Z, Sun J, Swale T, Yip HY, Qian PY, Qiu J-W, Kwan HS. 2020. Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing. *Nature communications* **11**:3051.

Noro Y, Shimizu H, Mineta K, Gojobori T. 2021. A single neuron subset governs a single coactive neuron circuit in Hydra vulgaris, representing a possible ancestral feature of neural evolution. *Scientific Reports* **11**:10828.

Noro Y, Yum S, Nishimiya-Fujisawa C, Busse C, Shimizu H, Mineta K, Zhang X, Holstein TW, David CN, Gojobori T, Fujisawa T. 2019. Regionalized nervous system in Hydra and the mechanism of its development. *Gene Expression Patterns* **31**:42–59. doi:10.1016/j.gep.2019.01.003

Orvis J, Albertin CB, Shrestha P, Chen S, Zheng M, Rodriguez CJ, Tallon LJ, Mahurkar A, Zimin AV, Kim M, Liu K, Kandel ER, Fraser CM, Sossin W, Abrams TW. 2022. The evolution of synaptic and cognitive capacity: Insights from the nervous system transcriptome of *Aplysia*. *Proc Natl Acad Sci USA* **119**:e2122301119. doi:10.1073/pnas.2122301119

Özbek S. 2011. The cnidarian nematocyst: a miniature extracellular matrix within a secretory vesicle. *Protoplasma* **248**:635–640. doi:10.1007/s00709-010-0219-4

Passano LM, McCullough CB. 1964. Co-ordinating systems and behavior in Hydra. I. Pacemaker system of the periodic contractions. *Journal of Experimental Biology* **41**: 643-664

Patel T, Hobert O. 2017. Coordinated control of terminal differentiation and restriction of cellular plasticity. *eLife* **6**:e24100. doi:10.7554/eLife.24100

Philipp I, Aufschnaiter R, Özbek S, Pontasch S, Jenewein M, Watanabe H, Rentzsch F, Holstein TW, Hobmayer B. 2009. Wnt/β-Catenin and noncanonical Wnt signaling interact in tissue evagination in the simple eumetazoan Hydra. *Proceedings of the National Academy of Sciences* **106**:4290–4295. doi:10.1073/pnas.0812847106

Pierobon P. 2015. Regional modulation of the response to glutathione in *Hydra vulgaris* (Pallas, 1766). *Journal of Experimental Biology* jeb.120311. doi:10.1242/jeb.120311

Piraino S, Boero F, Aeschbach B, Schmid V. 1996. Reversing the life cycle: medusae transforming into polyps and cell transdifferentiation in Turritopsis nutricula (Cnidaria, Hydrozoa). *The Biological Bulletin* **190**:302–312.

Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glažar P, Obermayer B, Theis FJ, Kocks C, Rajewsky N. 2018. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**:eaaq1723. doi:10.1126/science.aaq1723

Portugues R, Feierstein CE, Engert F, Orger MB. 2014. Whole-Brain Activity Maps Reveal Stereotyped, Distributed Networks for Visuomotor Behavior. *Neuron* **81**:1328–1343. doi:10.1016/j.neuron.2014.01.019

Prevedel R, Yoon Y-G, Hoffmann M, Pak N, Wetzstein G, Kato S, Schrödel T, Raskar R, Zimmer M, Boyden ES. 2014. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature methods* **11**:727–730.

Primack AS, Cazet JF, Little HM, Mühlbauer S, Cox BD, David CN, Farrell JA, Juliano CE. 2023. Differentiation trajectories of the Hydra nervous system reveal transcriptional regulators of neuronal fate. *bioRxiv*.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. 2007. Sea Anemone Genome

Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science* **317**:86–94. doi:10.1126/science.1139158
Rahat A, Rahat M, Searle JB. 1985. A simple method for the preparation of hydra chromosome spreads: introducing chromosome counts into hydra taxonomy. *Experientia* **41**:282–283. doi:10.1007/BF02002638
Raizen DM, Zimmerman JE, Maycock MH, Ta UD, You Y, Sundaram MV, Pack AI. 2008. Lethargus is a Caenorhabditis elegans sleep-like state. *Nature* **451**:569–572.
Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature communications* **9**:189.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. doi:10.1093/nar/gkw257
Rathje K, Mortzfeld B, Hoeppner MP, Taubenheim J, Bosch TCG, Klimovich A. 2020. Dynamic interactions within the host-associated microbiota cause tumor formation in the basal metazoan Hydra. *PLOS Pathogens* **16**:e1008375. doi:10.1371/journal.ppat.1008375
Reddien PW. 2018. The Cellular and Molecular Basis for Planarian Regeneration. *Cell* **175**:327–345. doi:10.1016/j.cell.2018.09.021
Reddy PC, Gungi A, Ubhe S, Galande S. 2020. Epigenomic landscape of enhancer elements during Hydra head organizer formation. *Epigenetics & Chromatin* **13**:1–16.
Reddy PC, Gungi A, Unni M. 2019. Cellular and Molecular Mechanisms of Hydra Regeneration. *Results Probl Cell Differ* **68**:259–290. doi:10.1007/978-3-030-23459-1_12
Reilly MB, Cros C, Varol E, Yemini E, Hobert O. 2020. Unique homeobox codes delineate all the neuron classes of C. elegans. *Nature* **584**:595–601. doi:10.1038/s41586-020-2618-9
Richards GS, Rentzsch F. 2015. Regulation of Nematostella neural progenitors by SoxB, Notch and bHLH genes. *Development* **142**:3332–3342. doi:10.1242/dev.123745
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**:460. doi:10.1186/s12859-018-2485-7
Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. 2018. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *cels* **6**:256-258.e1. doi:10.1016/j.cels.2018.01.001
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. doi:10.1093/bioinformatics/btp616
Rogers N, Cheah P-S, Szarek E, Banerjee K, Schwartz J, Thomas P. 2013. Expression of the murine transcription factor SOX3 during embryonic and adult neurogenesis. *Gene Expression Patterns* **13**:240–248. doi:10.1016/j.gep.2013.04.004
Ross KG, Currie KW, Pearson BJ, Zayas RM. 2017. Nervous system development and regeneration in freshwater planarians. *Wiley Interdisciplinary Reviews: Developmental Biology* **6**:e266.
Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily conserved principles predict 3D chromatin organization. *Molecular cell* **67**:837–852.

Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, Eliceiri KW. 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**:529. doi:10.1186/s12859-017-1934-z

Rushforth NB. 1965. Inhibition of contraction responses of Hydra. *American zoologist* **5**:505–513.

Rushforth NB, Burke DS. 1971. Behavioral and electrophysiological studies of Hydra. II. Pacemaker activity of isolated tentacles. *The Biological Bulletin* **140**:502–519.

Rushforth NB, Burnett AL, Maynard R. 1963. Behavior in Hydra: contraction responses of Hydra pirardi to mechanical and light stimuli. *Science* **139**:760–761.

Sachkova MY, Nordmann E-L, Soto-Àngel JJ, Meeda Y, Górski B, Naumann B, Dondorp D, Chatzigeorgiou M, Kittelmann M, Burkhardt P. 2021. Neuropeptide repertoire and 3D anatomy of the ctenophore nervous system. *Current Biology* **31**:5274-5285.e6. doi:10.1016/j.cub.2021.09.005

Sarras Jr MP. 2019. Hydra as a unique model for the study of regenerative mechanisms in metazoans. **Volume 6**. doi:10.15406/mojap.2019.06.00264

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**:495–502. doi:10.1038/nbt.3192

Schafer DP, Lehrman EK, Kautzman AG, Koyama R, Mardinly AR, Yamasaki R, Ransohoff RM, Greenberg ME, Barres BA, Stevens B. 2012. Microglia Sculpt Postnatal Neural Circuits in an Activity and Complement-Dependent Manner. *Neuron* **74**:691–705. doi:10.1016/j.neuron.2012.03.026

Schaible R, Scheuerlein A, Dańko MJ, Gampe J, Martínez DE, Vaupel JW. 2015. Constant mortality and fertility over age in Hydra. *Proceedings of the National Academy of Sciences* **112**:15701–15706. doi:10.1073/pnas.1521002112

Schenkelaars Q, Perez-Cortes D, Perruchoud C, Galliot B. 2020. The polymorphism of Hydra microsatellite sequences provides strain-specific signatures. *PloS one* **15**:e0230547.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**:676–682. doi:10.1038/nmeth.2019

Schwaiger M, Schönauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D, Technau U. 2014. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res* **24**:639–650. doi:10.1101/gr.162529.113

Schwentner M, Bosch TC. 2015. Revisiting the age, evolutionary history and species level diversity of the genus Hydra (Cnidaria: Hydrozoa). *Molecular Phylogenetics and Evolution* **91**:41–55.

Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé M-P, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S, Novault S, Steinmetz PRH, Spitz F, Tanay A, Marlow H. 2018. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **173**:1520-1534.e20. doi:10.1016/j.cell.2018.05.019

Seybold A, Salvenmoser W, Hobmayer B. 2016. Sequential development of apical-basal and planar polarities in aggregating epitheliomuscular cells of Hydra. *Developmental Biology* **412**:148–159. doi:10.1016/j.ydbio.2016.02.022

Shimizu H, Fujisawa T. 2003. Peduncle of Hydra and the heart of higher organisms share a common ancestral origin. *genesis* **36**:182–186.

Siebert S. 2018. Hydra cell lineages. http://commons.wikimedia.org/wiki/File:Hydra_cell_lineages.svg.
Siebert S, Anton-Erxleben F, Bosch TCG. 2008. Cell type complexity in the basal metazoan Hydra is maintained by both stem cell based mechanisms and transdifferentiation. *Developmental Biology* **313**:13–24. doi:10.1016/j.ydbio.2007.09.007
Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* **365**:eaav9314.
Simakov O, Bredeson J, Berkoff K, Marletaz F, Mitros T, Schultz DT, O'Connell BL, Dear P, Martinez DE, Steele RE. 2022. Deeply conserved synteny and the evolution of metazoan chromosomes. *Science advances* **8**:eabi5884.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. doi:10.1093/bioinformatics/btv351
Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* **6**:1–11.
Smith CL, Mayorova TD. 2019. Insights into the evolution of digestive systems from studies of Trichoplax adhaerens. *Cell and tissue research* **377**:353–367.
Smith K, Gee L, Bode H. 2000. HyAlx, an aristaless-related gene, is involved in tentacle formation in Hydra. *Development (Cambridge, England)* **127**:4743–52. doi:10.1242/dev.127.22.4743
Smith KM, Gee L, Blitz IL, Bode HR. 1999. CnOtx, a Member of the Otx Gene Family, Has a Role in Cell Movement in Hydra. *Developmental Biology* **212**:392–404. doi:10.1006/dbio.1999.9337
Sogabe S, Hatleberg WL, Kocot KM, Say TE, Stoupin D, Roper KE, Fernandez-Valverde SL, Degnan SM, Degnan BM. 2019. Pluripotency and the origin of animal multicellularity. *Nature* **570**:519–522.
Soriano J, Rüdiger S, Pullarkat P, Ott A. 2009. Mechanogenetic coupling of Hydra symmetry breaking and driven Turing instability model. *Biophysical journal* **96**:1649–1660.
Stover NA, Steele RE. 2001. Trans-spliced leader addition to mRNAs in a cnidarian. *Proceedings of the National Academy of Sciences* **98**:5693–5698. doi:10.1073/pnas.101049998
Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**:1888–1902.
Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**:15545–15550. doi:10.1073/pnas.0506580102
Sugiyama T, Fujisawa T. 1978. Genetic analysis of developmental mechanisms in Hydra. II. Isolation and characterization of an interstitial cell-deficient strain. *Journal of cell science* **29**:35–52.
Sun S, White RR, Fischer KE, Zhang Z, Austad SN, Vijg J. 2020. Inducible aging in Hydra oligactis implicates sexual reproduction, loss of stem cells, and genome maintenance as major pathways. *GeroScience* **42**:1119–1132. doi:10.1007/s11357-020-00214-z

Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* **5**:eaaw1668. doi:10.1126/sciadv.aaw1668

Szymanski JR, Yuste R. 2019. Mapping the Whole-Body Muscle Activity of Hydra vulgaris. *Current Biology* **29**:1807-1817.e3. doi:10.1016/j.cub.2019.05.012

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. 1988. Embryonic ε and γ globin genes of a prosimian primate (Galago crassicaudatus): Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *Journal of molecular biology* **203**:439–455.

Takahashi T. 2021. Hym-176 Handbook of Hormones. Elsevier. pp. 867–868.

Takahashi T, Hatta M, Yum S, Gee L, Ohtani M, Fujisawa T, Bode HR. 2005. Hym-301, a novel peptide, regulates the number of tentacles formed in hydra. *Development* **132**:2225–2234. doi:10.1242/dev.01792

Takahashi T, Koizumi O, Ariura Y, Romanovitch A, Bosch TC, Kobayakawa Y, Mohri S, Bode HR, Yum S, Hatta M, Fujisawa T. 2000. A novel neuropeptide, Hym-355, positively regulates neuron differentiation in Hydra. *Development* **127**:997–1005. doi:10.1242/dev.127.5.997

Takaku Y, Hwang JS, Wolf A, Böttger A, Shimizu H, David CN, Gojobori T. 2014. Innexin gap junctions in nerve cells coordinate spontaneous contractile behavior in Hydra polyps. *Scientific reports* **4**:3573.

Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, Wang B. 2021. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife* **10**:e66747.

Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, Poff A, Basavaraju M, Rafi I, Yemini E, Cook SJ, Abrams A, Vidal B, Cros C, Tavazoie S, Sestan N, Hammarlund M, Hobert O, Miller DM. 2021. Molecular topography of an entire nervous system. *Cell* **184**:4329-4347.e23. doi:10.1016/j.cell.2021.06.023

Technau U, Bode H. 1999. HyBra1, a Brachyury homologue, acts during head formation in Hydra. *Development (Cambridge, England)* **126**:999–1010. doi:10.1242/dev.126.5.999

Technau U, Rudd S, Maxwell P, Gordon PMK, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, Ball EE, Miller DJ. 2005. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends in Genetics* **21**:633–639. doi:10.1016/j.tig.2005.09.007

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **49**:D480–D489. doi:10.1093/nar/gkaa1100

Thomsen S, Bosch TCG. 2006. Foot differentiation and genomic plasticity in Hydra: lessons from the PPOD gene family. *Dev Genes Evol* **216**:57–68. doi:10.1007/s00427-005-0032-9

Tinevez J-Y, Perry N, Schindelin J, Hoopes GM, Reynolds GD, Laplantine E, Bednarek SY, Shorte SL, Eliceiri KW. 2017. TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**:80–90.

Tournière O, Dolan D, Richards GS, Sunagar K, Columbus-Shenkar YY, Moran Y, Rentzsch F. 2020. NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian Nematostella vectensis. *Cell reports* **30**:4473–4489.

Trembley Abraham, Pronk Cornelis, Schley Jacobus van der, Lyonet Pierre. 1744. Mémoires pour servir à l'histoire d'un genre de polypes d'eau douce, à bras en forme de cornes. A Leide, Chez Jean & Herman Verbeek.

Tursch A, Bartsch N, Holstein TW. 2020. MAPK signaling links the injury response to Wnt-regulated patterning in Hydra regeneration. *BioRxiv* 2020–07.
Uhlhaas P, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolić D, Singer W. 2009. Neural synchrony in cortical networks: history, concept and current status. *Frontiers in integrative neuroscience* 17.
Vaidya AR, Pujara MS, Petrides M, Murray EA, Fellows LK. 2019. Lesion studies in contemporary neuroscience. *Trends Cogn Sci* **23**:653–671. doi:10.1016/j.tics.2019.05.009
Vandestadt C, Vanwalleghem GC, Khabooshan MA, Douek AM, Castillo HA, Li M, Schulze K, Don E, Stamatis S-A, Ratnadiwakara M, Änkö M-L, Scott EK, Kaslin J. 2021. RNA-induced inflammation and migration of precursor neurons initiates neuronal circuit regeneration in zebrafish. *Developmental Cell* **56**:2364-2380.e8. doi:10.1016/j.devcel.2021.07.021
Vieira WA, Wells KM, McCusker CD. 2019. Advancements to the Axolotl Model for Regeneration and Aging. *Gerontology* **66**:212–222. doi:10.1159/000504294
Vladimirov N, Mu Y, Kawashima T, Bennett DV, Yang C-T, Looger LL, Keller PJ, Freeman J, Ahrens MB. 2014. Light-sheet functional imaging in fictively behaving zebrafish. *Nature methods* **11**:883–884.
Vogg MC, Beccari L, Ollé LI, Rampon C, Vriz S, Perruchoud C, Wenger Y, Galliot B. 2019a. An evolutionarily-conserved Wnt3/β-catenin/Sp5 feedback loop restricts head organizer activity in Hydra. *Nature Communications* **10**:312. doi:10.1038/s41467-018-08242-2
Vogg MC, Ferenc J, Buzgariu WC, Perruchoud C, Sanchez PGL, Beccari L, Nuninger C, Le Cras Y, Delucinge-Vivier C, Papasaikas P. 2022. The transcription factor Zic4 promotes tentacle formation and prevents epithelial transdifferentiation in Hydra. *Science Advances* **8**:eabo0694.
Vogg MC, Galliot B, Tsiairis CD. 2019b. Model systems for regeneration: *Hydra*. *Development* **146**:dev177212. doi:10.1242/dev.177212
Voigt O, Erpenbeck D, Wörheide G. 2008. A fragmented metazoan organellar genome: the two mitochondrial chromosomes of Hydra magnipapillata. *BMC Genomics* **9**:350. doi:10.1186/1471-2164-9-350
Voleti V, Patel KB, Li W, Perez Campos C, Bharadwaj S, Yu H, Ford C, Casper MJ, Yan RW, Liang W. 2019. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nature methods* **16**:1054–1062.
Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**:981–987. doi:10.1126/science.aar4362
Wagner G. 1905. Memoirs: on some movements and reactions of Hydra. *Journal of Cell Science* **2**:585–622.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963. doi:10.1371/journal.pone.0112963
Waltman L, Van Eck NJ. 2013. A smart local moving algorithm for large-scale modularity-based community detection. *The European physical journal B* **86**:1–14.
Wang H, Swore J, Sharma S, Szymanski J, Yuste R, Daniel T, Regnier M, Bosma M, Fairhall AL. 2020. From neuron to muscle to movement: a complete biomechanical model of Hydra contractile behaviors. *bioRxiv* 2020–12.

Watanabe H, Schmidt HA, Kuhn A, Höger SK, Kocagöz Y, Laumann-Lipp N, Özbek S, Holstein TW. 2014. Nodal signalling determines biradial asymmetry in Hydra. *Nature* **515**:112–115. doi:10.1038/nature13666

Weimer BR. 1928. The physiological gradients in Hydra I. Reconstitution and budding in relation to length of piece and body level in Pelmatohydra oligactis. *Physiological Zoology* **1**:183–230.

Weinziger R, Salgado LM, David CN, Bosch TCG. 1994. *Ks1* , an epithelial cell-specific gene, responds to early signals of head formation in *Hydra*. *Development* **120**:2511–2517. doi:10.1242/dev.120.9.2511

Weismann A. 1883. Die entstehung der sexualzellen bei den hydromedusen : zugleich ein Betrag zur Kenntniss des Baues und der Lebenserscheinungen dieser Gruppe / von August Weismann. Jena: Fischer.

Wells KM, Kelley K, Baumel M, Vieira WA, McCusker CD. 2021. Neural control of growth and size in the axolotl limb regenerate. *eLife* **10**:e68584. doi:10.7554/eLife.68584

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* **4**:1686. doi:10.21105/joss.01686

Wicks SR, Roehrig CJ, Rankin CH. 1996. A dynamic network simulation of the nematode tap withdrawal circuit: predictions concerning synaptic function using behavioral criteria. *Journal of Neuroscience* **16**:4017–4031.

Wilson-Sanders SE. 2011. Invertebrate Models for Biomedical Research, Testing, and Education. *ILAR Journal* **52**:126–152. doi:10.1093/ilar.52.2.126

Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG. 2006. Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proceedings of the National Academy of Sciences* **103**:6208–6211. doi:10.1073/pnas.0510163103

Wong WY, Simakov O, Bridge DM, Cartwright P, Bellantuono AJ, Kuhn A, Holstein TW, David CN, Steele RE, Martínez DE. 2019. Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus Hydra. *Proceedings of the National Academy of Sciences* **116**:22915–22917. doi:10.1073/pnas.1910106116

Yeo S, Coombe L, Warren RL, Chu J, Birol I. 2018. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* **34**:725–731. doi:10.1093/bioinformatics/btx675

Yoshida K, Fujisawa T, Hwang JS, Ikeo K, Gojobori T. 2006. Degeneration after sexual differentiation in hydra and its relevance to the evolution of aging. *Gene* **385**:64–70.

Yu S-M, Westfall JA, Dunne JF. 1985. Light and electron microscopic localization of a monoclonal antibody in neurons in situ in the head region of Hydra. *Journal of Morphology* **184**:183–193. doi:10.1002/jmor.1051840208

Yum S, Takahashi T, Koizumi O, Ariura Y, Kobayakawa Y, Mohri S, Fujisawa T. 1998. A Novel Neuropeptide, Hym-176, Induces Contraction of the Ectodermal Muscle in Hydra. *Biochemical and Biophysical Research Communications* **248**:584–590. doi:10.1006/bbrc.1998.8831

Zacharias H, Anokhin B, Khalturin K, Bosch TCG. 2004. Genome sizes and chromosomes in the basal metazoan Hydra. *Zoology* **107**:219–227. doi:10.1016/j.zool.2004.04.005

Zambusi A, Ninkovic J. 2020. Regeneration of the central nervous system-principles from brain regeneration in adult zebrafish. *World journal of stem cells* **12**:8.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**:R137. doi:10.1186/gb-2008-9-9-r137

Zhdanova IV, Wang SY, Leclair OU, Danilova NP. 2001. Melatonin promotes sleep-like state in zebrafish. *Brain research* **903**:263–268.

Zheng H, Xie W. 2019. The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology* **20**:535–550.

Zimmerman JE, Naidoo N, Raizen DM, Pack AI. 2008. Conservation of sleep: insights from non-mammalian model systems. *Trends in Neurosciences* **31**:371–376. doi:10.1016/j.tins.2008.05.001

Zimmermann B, Robb SMC, Genikhovich G, Fropf WJ, Weilguny L, He S, Chen S, Lovegrove-Walsh J, Hill EM, Chen C-Y, Ragkousi K, Praher D, Fredman D, Moran Y, Gibson MC, Technau U. 2022. Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. doi:10.1101/2020.10.30.359448