Science
AAAS

Supplementary Materials for

Profiling cellular diversity in sponges informs animal cell type and nervous system evolution

Jacob M. Musser *et al.*

Corresponding authors: Jacob M. Musser, jmmusser@gmail.com; Leonid L. Moroz, moroz@whitney.ufl.edu; Detlev Arendt, detlev.arendt@embl.de

Science **374**, 717 (2021)
DOI: 10.1126/science.abj2949

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Collection and cultivation of sponges

During winter, *Spongilla lacustris* individuals form protected packets of stem cells called gemmules, which can be collected and stored in the lab, and induced to hatch and form juvenile sponges. Adult specimens of *S. lacustris* in overwintering stage were collected on 12 January 2017 and 8 March 2018 from Lake Constance, near Kressbronn, Germany (47°35'09.0"N 9°35'56.4"E), compliant with local regulations. Individual sponge patches, composed of the sponge skeleton and gemmules, were stored in lake water at 4°C until gemmule isolation. Gemmules were extracted from the adult tissue via gentle agitation of the skeleton structure by lightly rubbing the sponge over sandpaper (grit 500) using one finger covered in nitrile gloves. Gemmules from different sponge patches were washed several times with either Vittel mineral water or M-medium (1 mM CaCl2·6H2O, 0.5 mM MgSO4·7H2O, 0.5 mM NaHCO3, 0.05 mM KCl, 0.25 mM Na2SiO3) (*38*) and stored separately at 4°C in Vittel mineral water.

Juvenile sponges were grown by placing gemmules in culture dishes with M-medium, and kept at room temperature (20-25°C) in the dark. To culture sponges for 10x single-cell RNAseq and smFISH, 4-8 gemmules were placed in a glass-bottom culture dish (Greiner Bio-One cat# 627860) containing 5 ml of Vittel mineral water. To culture sponges for Immunostaining, 4-8 gemmules were placed on a coverslip (24x24 mm, #1) inside 6-well plates, each well containing 5 ml of M-medium. To culture sponges for the contraction experiments, ~4 gemmules were placed on a coverslip (24x32 mm, #1) inside a 60 mm culture dish containing 10 ml of M-medium.

M-medium was replaced every other day, beginning around day 5 when juvenile sponges first adhere to the cover slip. Juvenile sponges were sampled or fixed for experiments after 7-12 days of growth, when they had acquired all major features of adults, including an osculum, well-developed canal system, and numerous choanocyte chambers. Variance in the exact time at which this stage occurs is extremely high, even for neighboring sponges grown from clonal gemmules that are present in the same culture dish. In general we tended to err on the side of letting sponges develop further, so that all sponges in the dish exhibited these features.

Bulk RNA isolation and sequencing for de novo transcriptome assembly

RNA was extracted from juvenile *S. lacustris* using Trizol (Thermofisher #15596026), and the resulting extracts were quantified using a Nanodrop 2100 and quality controlled on a Bioanalyzer 2100. Using strand-specific sequencing, we conducted 100PE and 125PE on a HiSeq 1500 using an RNA extract with RIN 10, generating 10,459,248 and 12,198,962 read pairs for the two samples, respectively.

Trinity de-novo transcriptome

Bulk RNA-seq reads were adapter trimmed, quality filtered, and assembled using Trinity (*39*), with the in-silico normalization option selected and other parameters set as default. This resulted in a transcriptome N50 of 905 base pairs. A BUSCO (*40*) search using the metazoa dataset indicated that the reference transcriptome was fairly complete (95.2% BUSCOs found). We identified putative proteins using Transdecoder (version 3.0.1), requiring a minimum open reading frame length of 100 amino acids.

Phylome, orthology inference, and de-novo transcriptome refinement

To infer orthology relationships between genes in the *S. lacustris* de-novo transcriptome with genes of other animals, we generated a comprehensive phylome containing trees for all genes in the *Spongilla* transcriptome. Briefly, each inferred *S. lacustris* protein was used as a seed to construct a gene/protein tree. For each seed, identification of the 250 closest homologs in 53 proteomes from other metazoan and unicellular relatives (listed in Data S1) was performed using BLAST with e-value and query coverage cutoffs of $E10^{-3}$ and 33%, respectively. Our choice of 250 homologs as the inclusion threshold came from our observation that this nearly always included the immediate orthology group, as well as several related paralogs that could serve as outgroups for orthology inference. It is important to note that our goal for the phylome pipeline was strictly to provide rigorous orthology inference for a given protein, and not to reconstruct the entire tree for large gene families. Phylogenetic workflow was executed using the workflow template named "clustalo_default-trimal001-none-raxml_default" from the ETE Toolkit v3.1.1 (*41*), which consisted of multiple sequence alignment reconstruction using clustalOmega v1.2.1 (*42*) with default parameters, TrimAl (*43*) for alignment cleaning by removing columns with more than 1% gaps, and Maximum Likelihood tree inference using RAXML v8.1.20 (*44*) with default parameters.

To validate our tree pipeline and ensure it produced robust results, we compared our automated trees to those published in previous studies where available. For example, we found congruence with previous studies of the *wnt* family (recovering 3 sponge ortholog groups) (*45*); *musashi1/2* (recovering a single demosponge gene and two calcareous sponge genes that are orthologous to *musashi1/2*) (*46*); *attractin* (single *Spongilla* ortholog) (*47*); *alpha-catenin* (one-to-three sponge ortholog of Ctnna1/2/3) (*48*), and beta-catenin (one-to-one ortholog) (*48*). We also inspected a large number of trees (>500), looking at their overall topology and the domain structure of aligned orthologs. In most cases, our trees broadly reflected the topology of animal relationships, finding distinct clades for major animal groups (e.g. chordates, ecydosozoans, lophotrochozoans, cnidarians, porifera, etc). Further, we found domain structure was broadly congruent within orthologous groups. These observations suggested our automated pipeline produced robust phylogenetic results useful for determining the orthology of *Spongilla* proteins.

To infer orthology information for each protein, each tree was rooted to the farthest leaf from the *Spongilla* seed protein using the function 'get_farthest_oldest_leaf' from the ETE Toolkit and using the NCBI Taxonomy tree as a reference of established relative dating for the species included. Rooted trees were automatically processed using ETE's `get_descendant_evol_events` function to infer speciation and duplication nodes. Speciation nodes were used to infer fine-grained orthology relationships (one-to-one, one-to-many, many-to-many) between seed sequences and other species.

To enable the public to make use of the *Spongilla* phylome resource, we produced a website (http://spongilla.compgenomics.org/) at which all phylome trees can be searched and viewed. Trees are indexed such that all trees containing a gene of interest can be found and viewed. Trees containing a given *Spongilla* gene are searchable via protein ID, automated gene name, or manually-curated gene name. Trees containing genes from *H. sapiens*, *M. musculus*, *D. melanogaster*, and *C. elegans* can also be identified by searching for their official UniProt names. Taxa in the trees are color-coded by animal clade (specified in Data file S1) to facilitate easier viewing of tree structure. Each tree is also accompanied by a visualization of the sequence alignment, with grey blocks representing regions of conserved sequence.

We next refined the assignment of transcripts to genes in our de-novo transcriptome using information derived from our phylome. We identified *S. lacustris* genes whose transcripts

and proteins did not align to each other but mapped as one-to-one orthologs to the same gene in the closely related freshwater demosponge *Ephydatia muelleri*. Visual inspection of many individual cases suggested these *S. lacustris* Trinity genes were fragments (e.g., 5‘ and 3‘) of the same gene. In a few cases, these also represented genes that were incorrectly fused in the *Ephydatia muelleri* transcriptome, although this represented only a small minority of cases (<5%). Based on this, we grouped those trinity genes that we inferred to represent fragments of the same gene into a single merged gene. The final merged gene set was then used to construct a GTF file for mapping our single-cell 3‘ RNAseq data and quantifying UMI counts for each gene. The assignment of trinity genes to their final merged gene IDs, which are used in the single-cell expression matrix, is available in the file “spongilla_merged_genes_and_names.tsv” on the project’s GitLab repository. We also provide full *Spongilla* transcriptome and proteome fasta files in the repository, in which sequence headers include the protein/transcript ID, automated gene name, and manually-curated gene name.

In total, our refined transcriptome contained 39,562 genes. Of these we identified 14,197 protein-coding genes containing an open reading frame of at least 100 amino acids. Our automated annotation pipeline assigned names to 9,809 protein-coding genes.

Functional annotation

We assigned orthology information and gene ontology terms to each protein-coding gene in our final refined transcriptome. Automated gene names and orthology assignments, used in our GTF annotation file and uploaded single-cell dataset, were assigned using an automated pipeline. First priority was given to orthology assignments to human genes inferred in the *Spongilla* phylome. However, for some *S. lacustris* genes, a human ortholog(s) was not identified by the phylome. For these, names were assigned by mapping the longest protein for each gene to the eggNOG database (version 4.5.1) using eggNOG-mapper v1 (*49*). In a few remaining cases in which neither the phylome nor eggNOG-mapper assigned a name, we identified a potential ortholog by taking the best hit via a blastp search against human RefSeq using default parameters. All gene names assigned via this automated pipeline use a similar nomenclature, beginning with a gene ID (c###_g###), followed by either the orthology relationship to human genes inferred from the *Spongilla* phylome, or the name designed via emapper or blastp.

For all gene names used in the manuscript’s main figures, and for some of the supplementary figures, we manually-curated the gene name by visually inspecting the gene tree, and also often examined domain structure using the online sequence search tool for InterPro (*50*). Manually curated gene names in the manuscript do not begin with a gene ID, which distinguishes them from the automated names used in the data matrix and in some of the supplementary figures. Rather, manually curated gene names follow a standardized nomenclature based off of their orthology to human genes (or in a few cases where orthology was uncertain, named based on the domains they contain or they gene family they belong to). For instance, a sponge gene orthologous to human genes *Lamp1* and *Lamp2* is named Lamp1/2. In cases where there are multiple sponge genes that are co-orthologs of the same human gene(s), these are given alphabetic designators (e.g. *Lamp1/2 A* and *Lamp1/2 B*). Lastly, in cases where the gene is inferred to be a paralog of a human gene, and for which no human ortholog exists, the gene is appended with “like” (e.g. *Myo7-like*). Manually-curated gene names are included in the *Spongilla* transcriptome and proteome fasta files, as well as in the file

"spongilla_merged_genes_and_names.tsv", all of which can be found on the project's GitLab repository.

Homeobox and Klf gene trees

For Homeobox and KLF gene families, which have traditionally posed challenged for phylogenetic reconstruction, we supplemented our automated phylogeniess from the *Spongilla* phylome with individual gene trees that we constructed using a more exhaustive pipeline. First, a cluster of homologous sequences for each gene family was manually selected using BLAST searches based on e-value and score to include all relevant transcripts and known genes from reference species. For the Homeobox gene family, all sequences with score >= 100 were included. For KLF, 647 sequences with at least 75 score were selected. The software UPP (*51*), specifically designed to improve the alignment of short fragments, was then used with parameters -M1 -B100 to infer a multiple sequence alignment for each family. Finally, a phylogenetic tree was inferred using IQTree with parameters -m TEST -bb 1000 -bnni.

Sample collection and sequencing for 10x single-cell 3' RNAseq

Dissociation of 8-day old juvenile *S. lacustris* in tissue culture dishes was performed by replacing M-medium with calcium and magnesium-free dPBS. Sponges were gently agitated by repeated pipetting, causing relatively rapid dissociation of all cells. Once completely dissociated (~5-10 minutes), cells were passed through a 40 μm mesh filter twice and collected in a low-bind 1.5 ml Eppendorf tube. Cells were washed twice by spinning at 1000g followed by resuspension in a small volume of calcium and magnesium-free dPBS. Cell viability was assessed using a combination of propidium idiodide (Cat #P4170, Sigma) and fluorescein diacetate (Cat #F7378, Sigma) to ensure greater than 90% live cells were used for capture. We conducted 4 single-cell RNAseq capture experiments from 8-day old juvenile *S. lacustris* using the 10x Chromium Controller and Chromium Single Cell 3' Kit v2 (Cat #120237, 10x Genomics, USA).

cDNA synthesis and library construction were made according to manufacturer's recommendations. Post-library quality control was determined on the Qubit fluorometer (Cat# Q32866, ThermoFisher Scientific) with the Qubit dsDNA HS Assay Kit and 1:10 sample dilution was run on the Agilent 4200 TapeStation system (Cat#G2964AA, Agilent Technologies) with the High Sensitivity D1000 ScreenTape (Cat#5067-5584, Agilent Technologies) and High Sensitivity D1000 Reagents (Cat#5067-5585, Agilent Technologies). Post-library quantification was performed with Illumina Library Quantification Kit (Cat #KK4824, KAPA Biosystems).

Single-cell libraries were sequenced on an Illumina NextSeq500 using a 2x150 paired-end kits with the following sequencing read recommendation Number of Cycles: 26 cycles Read1 for cell barcode and UMI, 8 cycles I7 index for sample index and 98 cycles Read 2 for the transcript.

Demultiplexing and mapping 10x single-cell RNAseq data

Individual 10x sample libraries were demultiplexed using Cell Ranger Makefastq v2.1.1 with default settings. Reads for each sample were mapped and demultiplexed by cell barcode and UMI using Cell Ranger Count v2.1.1. Lastly, we constructed a final aggregated count matrix using Cell Ranger Aggregate, which subsampled reads to ensure equal sequencing depth between all four samples. We note that for library 4 we captured relatively few cells, like due to a lower cell density in the dissociation used for capture. However, we found that cells from this library still behaved similarly in all QC and downstream analyses, and did not require any special

data integration methods to cluster with cells in the other samples. Based on this, we chose to include library 4 rather than discarding it.

Single-cell RNAseq QC, clustering, and gene set analysis

We visually inspected the distribution of genes, UMIs, and % mitochondral genes across cells. Ultimately, we chose to use a relatively inclusive threshold, removing cells expressing fewer than 200 genes, resulting in a final count matrix of 10,106 cells. We also tested other criteria for removing cells, but did not find that more stringent cutoffs yielded significantly different clustering results.

We explored a variety of clustering methods, including both Seurat and Monocle. In general, we found Seurat produced clusters with a clear biological signal, and abundant marker genes for each clusters. Building on this, we chose to use a combination of high-resolution Louvain clustering, implemented in R v3.5.1 using Seurat v2.4 (*52*), combined with a custom script for merging similar clusters (see GitLab repository). For this, we first selected highly variable genes using the approach of Macosko et al. (*53*). The top ~1000 genes were then used to perform principal components analysis. To identify principal components with significant variation, we performed a jackstraw resampling test, visually inspected the elbowplot of principal component vs. principal component standard deviation, and examined the genes with high loading for each principal component. Based on this we selected the top 40 principal components for cell clustering. However, we also tested using as few as 20, and as most as 100, of the top principal components, which generally yielded similar clustering results.

Clustering was conducted using the Louvain algorithm in Seurat. We explored extensively how Louvain clustering worked under different parameters, and on different datasets (e.g., stricter thresholds for cell inclusion, data subsets, etc). In general, clustering results were highly robust to different clustering parameters, with the exception of the resolution parameter, which significantly affected how many clusters we recovered. We found that clustering with higher resolutions was necessary to recover several rare but distinct cell types. However, this resolution setting also subdivided more common cell clusters, such as archaeocytes, into many smaller clusters that did not exhibit unique marker genes.

To identify all distinct clusters in our dataset, we opted for high-resolution clustering followed by iteratively merging nearest-neighbor clusters based on a similarity threshold. Briefly, clustering was first performed with the resolution parameter set to 10. We then calculated average expression profiles for each cluster, and used normalized expression vectors to calculate pearson correlations between each pairwise cluster combination. These were then ranked from highest to lowest correlation, and a Wilcoxon rank-sum test was used to calculate the number of differentially expressed genes between each pair. Cluster pairs that differed by less than 20 differentially expressed genes with 2-fold change were merged. This process was performed iteratively until all clusters were sufficiently distinct. Using this approach, we identified 42 genetically distinct clusters in our dataset.

We used the number of ‘specific’ markers to distinguish between clusters representing distinct cell types from those representing developmental transition states. ‘Specific’ markers for a cluster were defined as those cluster markers that were most highly expressed in that particular cluster (rather than in another cluster), as inferred using the Wilcoxon rank sum test. Clusters representing transitional states were designated as those clusters with less than 30 ‘specific’ markers, whereas those with more were designated as putative cell types. We found this threshold to be a conservative estimate of clusters representing distinct expression programs,

which was further consistent with our PAGA analysis (see below) that showed clusters with more than 30 specific markers often represented end-points along connectivity paths within the dataset. In total, we identified 23 putative cell types and 19 transitional states. Ultimately, we linked these 23 genetically distinct cell types to the 18 distinct morphological types referred to in the main text (with several morphological types exhibiting multiple semi-distinct expression states).

We identified co-expressed gene sets using weighted correlation network analysis, implemented in the R package WGCNA v1.68 (*54*). For this, we first identified an expanded set of ~6000 variable genes in Seurat to use in the analysis. We ran WGCNA on this pool of genes using mostly default parameters. Using the recommended approach, we set the softPower parameter to 5, required a minimum module size of 20, and conducted a final merging of gene sets whose eigengenes exhibited greater than 90% correlation.

PAGA connectivity graph

PAGA is implemented in Python, while the main analysis for this paper was conducted in R. To transfer the data between environments, we extracted the gene expression matrix after normalization, log-transformation, scaling, and centering, the cluster IDs after merging, as well as the PCA and tSNE coordinates, and used them to create an object we could analyse with the ScanPy package (*55*). We ran PAGA to calculate connectivity between the merged clusters, retaining all connections with a connectivity score above 0.7, which we found provided a good balance between preserving strong connections and removing those that were likely spurious. For visualization purposes, we colored strong edges (connectivity score of 1.0) black, and all other edges gray. Additionally, we used the connectivity value between clusters to determine the width of the edge connecting them, and plotted node size as the natural logarithm of the number of cells in the cluster. Finally, we placed the graph nodes on the median tSNE coordinates of the cells in the cluster, in order preserve the structure of the tSNE embedding and allow direct comparison between the PAGA graph and tSNE plot (note that the connections and connection strengths are not themselves inferred from the tSNE).

We also explored the robustness of the connections PAGA inferred among our clusters. Notably, we found that PAGA would occasionally create strong connections between genetically very dissimilar clusters. In particular, we found two of our small transitional clusters, 34 and 42, tended to form connections with many other clusters. Investigation of this showed that these connections were generated due to some cells being assigned to an apparently unrelated cluster (a phenomenon we have observed with Louvain clustering, and which may be caused by cells with high background or undetected doublets). These few cells effectively created a link between two otherwise unrelated clusters. To ensure our PAGA connections were not based on only a few cells with incorrect cluster assignments, we systematically inspected each cluster's cell assignments on our tSNE plot. We found clusters 34 and 42 had many cells scattered away from the main cluster, and falling next to other clusters in the tSNE plot. We also found a small number of cells in other clusters that exhibited a similar pattern. Rather than remove these cells, we chose a more conservative approach and simply reassigned these cells to the cluster with which they shared the highest pearson correlation (using each cluster's average expression profile to calculate the correlation). This soft cleanup of cell assignments removed several likely spurious connections. The remaining connections were generally supported by many cells with transitional identities and shared expression patterns, indicating they were more robust.

To investigate expression changes along different paths in the PAGA network, we calculated diffusion maps (*56*) and diffusion pseudotime as a proxy of the differentiation distance between each cell and the archaeocyte cluster. We then calculated the top marker genes for each PAGA path connecting archaeocytes with a distinct cell type cluster. For this we used the Wilcoxon rank-sum test, comparing each cluster on the path to all other cells on the path. Then, we plotted the expression of top marker genes for each cluster on the path (figs. S4-S10), or a select set of genes representing functionally relevant markers for choanocyte, myopeptidocyte, incurrent, and apendopinacocytes (Fig. 1E).

Lastly, although not reported here, we tested several other trajectory inference software pipelines, including exploratory analysis of the data with MERLoT (*57*). These results were largely consistent with what we observed using PAGA, providing further confirmation of the whole-data PAGA analysis.

Test for hierarchy in single-cell data and cell type tree construction

To evaluate the level of tree structure among genetically distinct cell types in our single-cell RNA-seq dataset, we first calculated the average value of log normalized counts per ten thousand for each putative cell type. We then calculated two types of distance matrices among cell clusters: the Euclidean distance for continuous expression values, and the Hamming distance for discretized values. We used a threshold of 0.025 for discretizing expression values, such that around 40% of the genes are marked as expressed in our analysis. Next, we followed the method in Liang, et. al. (*7*) to calculate the level of tree structure, π_0. Specifically, we calculated the δ-value for each group of four cell clusters (a tetrad) directly from the distance matrix. In total we estimated δ for a total of 8855 tetrads. δ values close to 0 indicate that the distance matrix of the tetrad conforms the requirement of a tree, whereas δ values close to 1 indicate that the tetrad lacks tree structure. We then estimate the significance of δ that comes from non-random gene expression using the formula: $p = \frac{3\sqrt{3}}{2\pi(\delta^2-\delta+1)}$. In this dataset, we obtained 23 different cell clusters and a total of 8855 tetrads. The Benjamini and Hochberg (BH) multiple test correction implemented in the 'qvalue' R package was utilized to estimate the fraction of non-random tree structure, π_0, which serves as an indicator of the level of tree structure in the dataset. Finally, the distance matrix was visualized using the SplitTree program (*58*).

To construct the *Spongilla* cell type tree, we calculated a distance matrix with the 'dist' function in R using average log normalized expression profiles for each of our 23 genetically distinct cell types, and restricting genes to the same ~6000 variable genes used in our WCNA gene set analysis. Neighbor-joining tree reconstruction was performed with the R package 'ape' v5.3. To test support for each node in the tree, we performed 10,000 bootstrap replicates. We also tested the robustness of our tree topology to the use of different gene sets. For this, we set two other thresholds for identifying variable genes. The first alternative gene set was composed of the top ~2500 variable genes, whereas the second contained the top ~12,000 variable genes. Neighbor-joining tree reconstructions showed that our cell type families and most intrafamily relationships were robust to the feature set used. Lastly, we rendered the final tree for the main figure using the Interative Tree of Life (iTOL) version 3 (*59*).

Cell type markers, family markers, and enriched gene ontology terms

To identify marker genes and enriched gene ontology terms for each cell type and cell type clade, we used two approaches. First, for each of the 23 clusters representing distinct cell type expression programs, we computed differentially expressed markers using the Wilcoxon rank-sum as implemented in the R package Seurat. For this, we computed markers by comparing cells within each distinct cluster, to all cells from the other distinct clusters (excluding cells from transitional clusters), and computing an adjusted p-value using the Benjamini-Hochberg (BH) procedure for controlling false discovery rate (*60*).

Second, we used evolutionary quantitative trait modeling to identify genes that change their expression in each clade in the tree. Specifically, for each gene, we used the mvMORPH package in R (*61*) to first fit a univariate Ornstein-Uhlenbeck model to the cell type tree that uses a single optimum value for the gene on the tree. This model represents the null expectation that the optimal expression value is unchanged across the tree. Then, for each clade in the tree, we fit a second model in which there is an additional optimal trait parameter for that particular clade. For each model fit, we provide the cell type tree and average expression value of a gene for each of the 23 distinct cell type clusters. Ornstein-Uhlenbeck model parameters are then fit using the mvOU function with default values, which uses a maximum likelihood estimation procedure to find the best parameter values. We then performed a likelihood ratio test using the LRT function in mvMORPH to determine whether the model with a second optimal trait parameter for a particular clade represented a better fit to the data than the model with only a single optimal trait value. To correct for multiple testing, we then adjusted the p-value using BH method for controlling the false discovery rate (*60*). Comparisons that were significant represented genes that had changed their expression values in a particular clade. We then applied this approach to all expressed genes, testing for changes in expression in each clade, and on all terminal branches. Results of this analysis are presented in Data file S1.

To determine functional enrichment of differentially expressed gene sets, or gene sets inferred to change their optimal expression value on the cell type tree, we conducted gene ontology enrichment using the GOseq package in R. For this we used default parameters values, but used the hypergeometric test to determine significance, as we found no evidence of sequence length-bias of gene enrichment in our 3' single-cell data (correcting for length-biased did not change results). We also compiled custom gene ontology terms for synapse and neural-related categories, for the purpose of testing their enrichment in different cell types or cell type clades. For these we drew extensively from the literature (*21, 22, 62–71*), allowing us to refine these gene ontology terms down to those genes that have been shown to play a central functional role in these processes. For each ontology term, we then tracked down all orthologs in *Spongilla*, and coding all of them into our *Spongilla* gene ontology file (available on the gitlab repository – custom GO terms include the phrase "manually curated" in the name to allow for easy identification). Enriched custom GO terms are found in Data S1 in our GO term enrichment analysis results (findable by searching for the phrase "manually curated").

SAMap alignment of *Spongilla* and adult *Amphimedon* single-cell datasets

We used SAMap (*8*), a method for comparing single-cell datasets between evolutionarily distant species that is robust to different types of single-cell data, to align the 10x *Spongilla* (10,106 cells) and MARS-seq adult *Amphimedon* (3,870 cells) (*9*) single-cell atlases. SAMap requires the raw counts from each dataset and a comprehensive list of their homologous gene pairs, which we identified using reciprocal tblastx with an e-value threshold of $1e^{-6}$. SAMap then uses a graph-stitching approach to link the datasets and subsequently refines the list of

homologous gene pairs to include those with overlapping expression patterns across the aligned manifold. This process repeats for a specified number of iterations or until the aligned manifold converges. We ran SAMap with the following parameters: NH1=2, NH2=2, and NUMITERS=6. NH1 and NH2 determine the number of hops away from each cell in dataset 1 and 2, respectively, to include in its neighborhood. SAMap averages the cross-species edges between cells across these neighborhoods to improve the robustness and sensitivity of the mapping. We set NH1=2 and NH2=2 as opposed to the default values of 3, as the SAMap tutorial vignette (https://github.com/atarashansky/self-assembling-manifold) recommends decreasing neighborhood sizes for small datasets to ensure that the neighborhoods of distinct cell types do not overlap. NUMITERS determines the number of iterations SAMap will run. We increased the number of iterations from the default value of 3 to 6 to allow ample time for the aligned manifold to converge.

The species-colored UMAP projection was generated using the scatter function in SAMap. The expression overlap UMAP plots were generated using the plot_expression_overlap function in SAMap. Expressing cells are color-coded by species, with those connected across species colored cyan. The pairs of genes displayed in Figure S13 were identified using the samap.analysis.GenePairFinder method in SAMap. This method identifies pairs of genes that are both expressed in connected cell types and differentially expressed in their respective datasets. The Sankey plot was generated using the sankeyD3 package in R. Alignment scores depicted in the Sankey plot range from 0 to 1 and represent the average number of mutual nearest cross-species neighbors of each cell in a cluster relative to the possible number of neighbors.

Single-molecule FISH

Prior to fixation, juvenile sponges growing in glass-bottom culture dishes were incubated with 200 µl of CellBrite Fix 488 membrane staining solution (stock diluted 1:1000 in Vittel mineral water or M-medium; Cat #30090, Biotium) for 15 min at RT, washed twice with Vittel water, and then fixed with 2% paraformaldehyde in ¼ Holtfreter's solution (¼ HS: 87.5 mg NaCl, 1.25 mg KCl, 2.5 mg CaCl2, 5 mg NaHCO3 in 100 mL H_20, Funayama et al. 2005) for 20 min at RT (room temperature). Specimens were washed once in ¼ HS for 5 min, then dehydrated with 50% MetOH in ¼ HS, 100% MetOH, and 100% EtOH. Specimens were then permeabilized with 200 µl of 50% Xylene/EtOH for 15 min at room temperature (RT), while trying to minimize the exposure of Xylene to the plastic sides of the dish as much as possible. Specimens were rehydrated in a series of 75%, 50% and 25% EtOH in PBS-Tr (PBS with 0.1% Triton-X100), and then washed three times in PBS-Tr for 10 min each. Sponge specimens were transferred to 1 ml Hybridization buffer (2x SSC pH 7.0, 15% Ethylene Carbonate, 1 mM EDTA, 50 µg/ml heparin, and 1% Triton X-100 and were pre-hybridized for at least 10 min at 37°C. Hybridization was carried out overnight at 37°C with 3 nM labeled ssDNA oligos (i.e. smFISH probes) in 200 µl Hybridization buffer. smFISH probes were constructed against target RNA following Gaspar et al. (*72*). DNA Oligos (Data S1) were bound to either ddUTP-Atto633 (Cat #25040, Lumiprobe and Cat #AD633, Atto Tec) or ddUTP-Atto565 (Cat #Ad565, Atto Tec) fluorophores. During hybridization, the dishes were covered with parafilm to prevent evaporation. After hybridization, samples were washed with 1 ml Hybridization buffer (without smFISH probe) three times for 15 min each at 37°C. Specimens were then washed three times with 3 ml PBS at RT. For DNA counterstaining, sponge samples were incubated with 1 ml DAPI staining solution (1 ug/ml DAPI in PBS; Cat #D9542, Sigma-Aldrich) for 20 min at RT and subsequently washed twice with 3 ml PBS. Sponge specimens were stored in PBS at 4°C for up to one week. Images were

taken with a 40x water immersion objective on a Leica SP8 inverted confocal microscope, and images were processed using Huygens deconvolution software and ImageJ (z-stack: maximal projection intensity projection).

Immunostaining, confocal, and STED imaging

Immunostaining procedure was performed as described previously (*73*). Primary antibody incubation was with Acetylated tubulin (1:500 dilution; Cat #T6793, Sigma-Aldrich), and secondary antibody incubation was with Alexa Fluor 488 Goat, Anti-Rabbit IgG (H+L; 1:500 dilution; Cat #111-545-144, Jackson ImmunoResearch), Alexa Fluor 568 Phalloidin (1:40 dilution; Cat #A12380, ThermoFisher), and DAPI (1 ug/ml). Images were taken with a 63x oil immersion objective on a Leica SP8 inverted confocal microscope, and images were processed using ImageJ (z-stack: maximal projection intensity projection) and Imaris x64 v9.3.1.

For Stimulated emission depletion (STED) microscopy, we used a rabbit polyclonal primary antibody against ezrin, moesin, and radixin protein (1:100 dilution; Cat # ab76247, Abcam) and a mouse monoclonal antibody against Acetylated tubulin (1:50 dilution; Cat #T6793, Sigma-Aldrich). Primary antibodies were coupled with the following secondary antibodies: goat anti-rabbit Star 580 (1:500 dilution; Cat# ST580; Abberior) and goat anti-mouse Alexa 488 (1:500 dilution; Cat# 111-545-144; Jackson ImmunoResearch). After antibody staining we incubated the samples with Phalloidin Atto-647N (1:40 dilution, Cat# 65906; Sigma) and Dapi (1:500 dilution), performed three final rinses in PBS, and mounted the samples in Mowiol. Samples were imaged on an Abberior STED/RESOLFT microscope (Abberior Instruments, Expert Line) running the Imspector software (Abberior Instruments). The microscope comprises an IX83 (Olympus) in combination with a UPlan-S Apochromat ×100/1.40 NA oil objective (Olympus). Super-resolved images were acquired by donut-shaped depletion using a 775 nm pulsed laser along with a 640 nm and 594 nm pulsed laser, exciting Phalloidin Atto 647N and Star 580 labelled ERM protein. Images were processed using ImageJ.

Western Blot (ProteinSimple Jess) of ERM antibody

Specificity of the ezrin, radixin, and moesin antibody to the single *Spongilla* ortholog was determined by Western Blot (fig. S21F). For this, whole cell lysates were derived from 8-day old juvenile sponges grown at 18 °C in the dark. 15 sponges were transferred to a 1.5 mL reaction cup and dissolved in lysis buffer (D-PBS, 0.5 % Triton X-100, 0.1 % SDS, 0.4 % Sodium deoxycholate, 1 mM EGTA, 1.5 mM MgCl2, 1 x protease inhibitor cocktail (cOmpleteTM, Roche), 0.26 U/µL benzonase (Millipore) by shaking for 1 h, 500 rpm at 4 °C. The lysate was clarified by centrifugation at 21.000 x g for 30 min at 4 °C. Validation of the ERM antibody was performed using an automated capillary-based immunoassay platform (ProteinSimple Jess). According to the manufacturer's instructions, a 12-230 kDa separation module (SM-W004) for anti-rabbit chemiluminescence detection (DM-001) was prepared. Commercially available anti-ERM (abcam, ab76247) antibody was used at a 1:10 dilution. The run was performed using default settings and analyzed using the Compass Software for Simple Western (version 5.0.1, ProteinSimple).

The expected weight of the *Spongilla lacustris* ERM ortholog is around 66 kDa. Comparable to the western blot results in HeLa cell lysates, the observed band size of 72 kDa is slightly larger than the expected weight. Similarly, the results reported using HeLa cell lysate show extra bands between 50 and 15 kDa. This finding, together with the results of the

immunostaining showing localization to microvilli, is consistent with the antibody binding to the *S. lacustris* ERM ortholog.

Imaging endogenous sponge contractions cycles

Time-lapse imaging for the general contraction model was performed as modified from Nickel (*74*) using a computer-controlled Nikon Coolpix 990 attached to a dissecting microscope (Zeiss SM XX or Zeiss Stemi 2000) through a custom made projection adapter (Zeiss GF-PK 10/20 ocular glued to a Kendo 28 mm adapter ring). The time-lapse was triggered through a PC running DC-Time-Trigger and DC-Remote-Shutter 2.0.1. software (Per Madsen, digitalkamera.dk). Illumination was achieved by cold-light source (Zeiss KL 200) set on low intensity in order to avoid heating up the sponges.

Sponges were imaged in 60 mm Petri dishes. For bi-planar views, a coverslip (24x24 mm, Roth) was gold-sputtered in a sputter coater (Emitech K500) to serve as a mirror, which was introduced at a 45° angle as close as possible to the imaged sponge. The mirror coverslip was precisely positioned using a 3-way micro-manipulator (Zeiss). The camera field of view was adjusted through the dissecting microscope zoom lens to fit both views simultaneously, the axial view as well as the mirror-based lateral view.

Image processing of endogenous contractions

Image analysis and time-lapse video creation was performed in ImageJ. For measurements in axial and lateral views, grey value thresholds were used to represent total projected area of lateral and axial projected areas, based on a custom-made loop macro, measuring a whole image stack. Projected areas of subdermal lacunae, incurrent canals, excurrent canals, and oscule were manually measured at given positions (oscule in lateral view) or partial areas (M1-M3 in axial and lateral views).

Flow-through chamber for NO experiments

We utilized a flow-through system to enable introducing and flushing away substances to juvenile sponges without directly touching the dish or chamber holding the sponge. This is critical, as mechanical vibration is known to stimulate sponge contractions. Juvenile sponges grown on coverslips were inserted into a micro observation chamber (*75*) of 1.7 ml volume which was placed on an inverted microscope (Zeiss Cell Observer). The chamber was part of a custom-made gravity-driven continuous flow system, with reservoir containing fresh M-medium placed on a shelf above the microscope, and connected to the chamber via flexible medical plastic tubing. A collection reservoir, also connected to the chamber via medical tubing, was placed below the chamber. Partial closure of the tubing enabled for precise control of the flow-rate through the chamber containing the sponges, which was typically set between 10-50 mL per hour. We built in luer lock adapters and 3-way-valves to allow for the introduction of experimental substances via injection with a 30 mL syringe.

Nitric oxide and ODQ experiments

We introduced NOC-12 (Cat #E3145, Sigma-Aldrich) or ODQ (Cat #O3636, Sigma-Aldrich) separately and in serial combinations to juvenile sponges in our flow-through chamber. We note that ODQ's inhibition of the soluble Guanalyl cyclase inhibitor has not been directly tested previously in sponges. However, experiments in a wide variety of other invertebrates indicate ODQ consistently acts in an antagonistic manner to NO treatment, and causes a decrease

in cGMP concentrations, consistent with the inhibition of the soluble guanalyl cyclase receptor. We also performed control experiments in which we halted flow-rate or introduced DMSO. For each experiment, juvenile sponges were introduced into the flow system 4-12 hours before injection, allowing them to equilibrate under a constant flow rate of 10-50 mL M-medium per hour. Following this, we gently injected 15 mL of the treatment solution, which ensured that the observation chamber and all tubing leading in and out of the chamber, up to the valves, were flushed with the treatment solution. For experiments with single treatment solutions, we alternated incubation periods with washing periods in which normal flow rates of M-medium were resumed. For experiments containing both NOC-12 and ODQ, we first incubated sponges with one substance, followed by incubation with both substances, before lastly washing by resuming the normal flow of M-medium.

Experiment	**Treatment solution**	**Concentration**	**Experimental sequence**
Nitric oxide	NOC-12	22.7 µM	Normal flow (12 hrs), NOC-12 (12 hrs), wash/normal flow (12 hrs), NOC-12 (24 hrs), wash/normal flow (12 hrs), NOC-12 (8 hrs), wash/normal flow (48 hrs)
sGC inhibition	ODQ	3 µM	Normal flow (26 hrs), ODQ (8 hrs), wash/normal flow (13 hrs), ODQ (12 hrs), wash/normal flow (12 hrs), ODQ (24 hrs), wash/normal flow (48 hrs)
Nitric oxide followed by sGC inhibition	Noc-12, ODQ	22.7 µM, 3 µM	Normal flow (21 hrs), NOC-12 (8 hrs), NOC-12/ODQ (16 hrs), wash/normal flow (12 hrs), NOC-12 (12 hrs), NOC-12/ODQ (24 hrs), wash/normal flow (11 hrs), NOC-12 (24 hrs), NOC-12/ODQ (16 hrs), wash/normal flow (48 hrs)
sGC inhibition followed by Nitric oxide	ODQ, Noc-12	3 µM, 22.7 µM	Normal flow (19 hrs), ODQ (2 hrs), ODQ/NOC-12 (11 hrs), wash/normal flow (11 hrs), ODQ (9 hrs), ODQ/NOC-12 (16 hrs), wash/normal flow (12 hrs), ODQ (17 hrs), ODQ/NOC-12 (27 hrs), wash/normal flow (48 hrs)
Control	DMSO		Normal flow (24 hrs), DMSO (8 hrs), wash/normal flow (12 hrs), DMSO (24 hrs), wash/normal flow (24 hrs)

Time-lapse imaging of nitric oxide experiments

During contraction experiments, images were taken using a consumer camera (Panasonic G2), attached to the microscope through a custom-made ocular adapter through a MFT-T2 adapter which included an 8x projective (Olympus). The camera was connected to a flash unit (Nikon SB24), covered by a red diffuser foil. Both, camera and flash were connected to a permanent power source (Panasonic DMW-BLB13 and custom made). Time-lapse imaging was triggered by an intervalometer set either to 30 or 60 sec intervals. Images were stored on the camera using a wifi-SDcard (ez Share), which was connected to a PC running ez Share Windows Client V1.1.0 under Microsoft Windows, transferring Images directly to a network drive after capturing.

Image processing of treated sponges

Images were processed in FIJI in order to remove background noise on the coverslip surrounding the sponge, which tends to accumulate over time. A semiautomated macro was used to segment the total sponge area: Based on the wand tool functionality the macro was provided with a pixel coordinate inside the sponge, plus a lower (depending on the single experiment, which slightly varied due to the sponges in use) and upper grey (255) value threshold, from which the threshold area was grown based on 8-connectivity. The resulting outline of the sponge was expanded by 10 pixels in all directions and the background was removed (set to 0/black). This was automated across the time stacks taken.

Segmenting images of whole sponges is challenging as they exhibit a complex three-dimensional topology. To infer relative sizes of mesohyl, choanoderm, and canal compartments, we used grey value thresholding (*12*), tuning thresholds independently for each sponge to best capture each compartment. Relative pixel measurements were obtained for the full time series in FIJI and plotted in R using 'ggplot2' v3.1.1.

FIB-SEM Pre-fixation and Fluorescence Confocal Microscopy

Spongilla were grown on Ibidi Glass Bottom Dishes 35 mm for 14 days and were prefixed with aldehydes and stained with DAPI. Due to the high osmotic sensitivity of these freshwater animals, we performed the pre-fixation in two steps: first for 30 min at room temperature in 4% FA in ¼ HS medium and subsequently in 2.5% GA and 1.6% FA in PHEM in a microwave (Pelco Biowave containing a ColdSpot cooling system, 150-W on/off cycling intervals of 7 x 2 min, under vacuum). PHEM was prepared by dilution of a 4x PHEM stock, containing PIPES 240 mM, HEPES 100 mM, $MgSO_4$ 8 mM, and EGTA 40 mM in ddH_2O and adjusted to pH 6.9 with KOH

Samples were then stained with DAPI (20 min, 2 µg/ml in PHEM), rinsed twice in PHEM and kept in the same buffer during confocal acquisition. We identified putative choano-neuroid cells based on the presence of DAPI staining in the center of a choanocyte chamber. Locations of target choanocyte chambers, hosting a central putative neuroid cell, were mapped in three dimensions using the DAPI fluorescence, imaged on a Leica SP8 confocal microscope (10x and 20x dry objectives). We acquired images of a large portion of the sponge bodies, focusing on the edges that are the most EM-accessible regions. Confocal acquisition permitted the accumulation of morphological landmarks later used to trace back the selected choanocyte chamber for the FIB-SEM acquisition.

Post-fixation for FIB-SEM

After confocal acquisition, samples were post-processed for FIB-SEM using an adaptation of the rOTO protocol (*76*, *77*) aided by microwave processing, as detailed by Schieber et al. (*78*). Post-fixation steps are summarized in Table 1. Durcupan was freshly prepared by mixing the components A-D in the following quantities: A: 11.4 g, B: 10 g, C: 0.3 g, and D: 0.1 g.

	Post processing steps	**Duration and incubation mode**
1.	1% OsO_4 and 1.5% $K_4Fe(CN)_6$ in PHEM	In microwave, 7 x 2 min at 150 W cycling on and off, vacuum on
2.	Rinse 2 times in PHEM	3 min each in the microwave
3.	Rinse 5 times in water	3 min each in the microwave
4.	1% TCH in water	10 min on the bench
5.	OsO4 1% in water	In microwave, 7 x 2 min cycling on and off at 150 W, vacuum on
6.	Rinse 2 times in water	3 min each rinse in the microwave
7.	1% UA in the microwave	7 x 1 min at 150 W cycling on and off in vacuum
8.	Dehydration series in ethanol (10%,25%,50%,75%,90%,95%,100%x3)	40 sec each in the microwave, vacuum off
9.	Infiltration series in Durcupan diluted in ethanol, repeat twice each step: 10%,30%,50%,70%,90%,100%	2 min each in the microwave, vacuum on
10.	Change to freshly prepared Durcupan resin for flat embedding (by leaving an about 2mm thick resin layer on the coverslip) and polymerize	72 hrs at 60°C in regular oven

The glass coverslips were detached from the resin surface using multiple heat shocks in liquid nitrogen and hot water. Resin blocks were then trimmed down to the size of the individual sponges, into 3-4 mm large and about 2 mm thick blocks.

Targeted trimming for volume EM of one full choanocyte chamber

Targeted trimming is necessary to expose the region of interest to FIB-SEM imaging, i.e. placing the targeted cell not deeper than 10 to 30 µm from the block surface, both in the imaging and in the milling directions. To achieve this step with precision, the topology of the resin-embedded sponges was characterized by comparing the view of the resin-embedded specimen with the confocal images acquired before embedding. With this information, the blocks were trimmed using a glass knife and an ultramicrotome (Leica EM UC7), creating a surface 10 µm above the target. Note that it is recommended not to use diamond knives on this sample, as Spongilla produces thick and hard silica spicules. The trimmed blocks were then mounted on FIB-SEM stubs using conductive epoxy resin, polymerized overnight at 60 ℃. Finally, a layer of silver paint and gold sputter coating were applied.

Targeting strategy for high-resolution imaging of neuroid cells

To establish the coordinates of individual neuroid cells within the resin block to an accuracy sufficient for FIB-SEM data acquisition of volumes containing an individual cell, we used a combination of microscopic X-ray computed tomography (microCT) and phase-contrast X-ray imaging on a synchrotron beamline (fig. S29, A to D). A 3D X-ray tomogram of the entire volume of a block containing the sample was acquired with a Bruker SkyScan 1272. Although the tomogram revealed a number of morphological features, the resolution was not sufficient to visualize neuroid cells.

For 3D X-ray tomographic imaging with synchrotron radiation (XIMG), the resin block was glued to a SPINE sample holder commonly used for macromolecular crystallography (*79*). Following standard procedures, the sample holder was inserted into a plastic vial which was then mounted into a protecting metal sample basket for standard courier-shipment from the Heidelberg laboratory to the synchrotron beamline in Hamburg.

XIMG data were collected on the EMBL beamline P14 on the PETRA III storage ring (c/o DESY, Hamburg, Germany) using the propagation-based phase-constrast imaging setup described in Polikaropv et al. (*80*) at an X-ray energy of 15.5 keV. X-ray images were recorded using an Optique Peter (Lyon, France) X-ray microscope consisting of an LSO:Tb scintillator with 8 μm active layer; an Olympus UPlanFL 20-fold objective (Olympus, Tokio, Japan), numerical aperture 0.5; a 45° mirror; a 180 mm tube lens and a PCO.edge 4.2 sCMOS camera with 2048x2048 pixels, pixels (6.5 μm pixel size). Thus, the effective pixel size was 0.325 μm with a field of view 666 x 666 μm². This setup typically delivers a resolution of about 0.5-0.7 μm, as determined from the analysis of projection images from a Siemens star (Ta on SiN; XRESO-50HC, NTT-AT, Japan). On the resin-embedded sponge sample, projection images were acquired at four camera distances: 136, 141, 146 and 151 mm as selected according to Zabler et al. (*81*). At each distance, 30 flat-field images and 1850 projections covering 185° of continuous rotation were recorded with an exposure time of 10 ms per frame. Data collection (including robotic sample transfer from a storage vessel to the rotation axis and sample centering via an on-axis optical microscope) was completed in 3.5 minutes.

Flat-field corrections were applied by dividing each projection image by the most similar flat-field image according to the SSIM criterion (*82*). For lateral shift compensation at the four camera positions, images recorded at each projection angle were registered using Fourier-space correlation with a sub-pixel interpolation. Registered images were further processed by a multi-distance non-iterative holographic reconstruction (*81, 83*), using a complex refraction index decrement ratio $\beta/\delta = 0.075$ and a zero compensation of 0.001. Tomographic reconstructions were performed using the TOMOPY package (*84*), employing the built-in Gridrec algorithm and Shepp-Logan filtering with default settings. All steps of the XIMG data processing were combined into a python-based custom software pipeline, available at: https://pypi.org/project/maximus48/. The tomogram acquired with synchrotron radiation revealed sufficient detail for the detection and accurate localization of neuroid cells in three dimensions (fig. S29B).

The XIMG-based tomogram was then registered into the coordinate system of the full block using Amira (Thermo Fischer Scientific), based on features visible in both datasets. A preliminary alignment with the editor was refined automatically using a 3D rigid Euclidean transformation in the Multiplanar viewer (fig. S29, A to D). We were thus able to accurately localize the cells of interest within the block's coordinate system. This information guided the trimming of the block surface, which was then done with an ultramicrotome (Leica UC7). Using

an iterative trimming process, and by comparing the predicted XIMG virtual slice with the current block face imaged at the SEM, we were able to reach a position approximately 11 µm 'above' the neuroid cell suitable for starting the FIB-SEM acquisition (fig. S29, E to F). By registering in 2D corresponding features from the XIMG virtual slices and the block surface, we could thus measure the exact X,Y location of the cell of interest (fig. S29, E to F), and use this information to define the position of the FIB-SEM trench (fig. S29G). Next, virtual XIMG slices perpendicular to the block surface were registered with the SEM image of the trench side, allowing us to place the acquisition window in the right position (fig. S29, H to J). The targeted volume contained the neuroid cell, as expected by the XIMG imaging (fig. S29, K to M).

Volume EM by FIB-SEM acquisition

We imaged the samples on a Zeiss Crossbeam 540, using Atlas 3D v5 (FIBICS, Carl Zeiss Microscopy) to guide the acquisition. The surface of the block was platinum-coated in correspondence with the predicted location of the target choanocyte chamber. For the acquisition of the full choanocyte chamber, FIB slicing during the run was obtained at 3nA. The dataset was acquired using the SEM at 1.5kV and 700pA current, using a back-scattered electron detector (ESB). Imaging occurred in three separate sessions for a cumulative time of about 96 hours, during which the acquisition area was continuously optimized in order to accommodate the full choanocyte chamber and minimize charging. The final acquired volume was 60 µm x 60 µm x 50 µm with an isotropic voxel resolution of 15 nm.

For the acquisition at a high resolution of the single neuroid cells, we exposed the imaging surface by FIB milling of a 30 µm deep trench from the resin block. FIB slicing was obtained at 700pA. SEM acquisition was obtained as described above, with a voxel size of 8x8x8 nm for cell 1 and 5x5x8 nm for cell 2.

FIB-SEM data pre-processing

The resulting image series was loaded as a Virtual stack in Fiji, cropped to the regions of interest and converted to 8-bit using batch processing. Alignment occurred in two steps: first by SIFT alignment in Fiji (image J ref) and then using a custom-written Python script for subpixel alignment.

Data visualization and rendering

A selection of cells, cilia and microvilli were segmented as detailed above, exported as .obj meshes into Blender 8.0 for visualization and rendering. For visualization of the whole volume, the pre-processed image sequence was binned 2 x 2 x 2, corrected for the differences in background between the different acquisition sessions and imported into Drishti v2 (*85*), using the drishtiimport.exe program.

To study the interaction of choanocytes and central cells in the EM volume, we produced an instance segmentation of the relevant structures, namely cell bodies, microvilli, and cilia of choanocytes, apopylar cells, and choano-neuroid cells. The size, shape, and frequency of objects of these three categories differ significantly. In addition, the intensity histogram shifts between parts of the data-set that were imaged in different sessions. These facts make the task at hand too challenging to be solved by a thresholding-based approach. Thus, we employed the Lifted Multicut (*86*) based segmentation workflow developed in Pape et al. (*87*) for instance segmentation problems in bio-medical images. This approach consists of two main steps: first, a machine learning approach is used to predict per voxel probabilities for object boundaries and

semantic classes. Based on these predictions, a graph-based segmentation problem is formulated, where boundary evidence is mapped to local interactions and semantic evidence to non-local interactions.

Here, we use the autocontext workflow of ilastik (*88*) for the first step. This algorithm consists of multiple stages of voxel classification. In an individual stage a Random Forest Classifier (*89*) predicts semantic class probabilities for each voxel. The classifier is trained from sparse labels with ilastik and uses the stacked responses of several convolutional filters applied to the input data as features. Each stage is presented with the raw data and the predictions of the previous stage; thus refining the previous predictions. We perform three stages of Autocontext: in the first and second stage, we predict six different classes: *background*, *object boundary* (corresponding to membranes of cells, flagella and microvilli), c*ytoplasm*, *nucleus*, *flagellum* and *microvillus*. In the third stage, we predict the binary classes *boundary* and *non-boundary*.

We use the boundary predictions from the third stage and the class predictions from the second stage to perform Lifted Multicut based segmentation. This approach involves several sub-steps: first, the boundary predictions are used to compute supervoxels via distance-transform watersheds (See *90* for details). These supervoxels are mapped to a graph which represents each supervoxel by a node. Two nodes are connected via an edge if the corresponding supervoxels are adjacent. In order to formulate the segmentation problem, we compute edge weights from the boundary predictions; these weights express the likelihood that the incident nodes belong to the same object. Following the procedure outlined in Pape et al. (*87*), we augment this graph by sparse lifted edges. To derive these edges, we first compute the connected component segmentation of the thresholded microvilli and flagella predictions. Then we map the resulting objects to graph nodes via overlap with the corresponding supervoxels and introduce lifted edges between nodes mapped to the same objects. We associate attractive weights with the lifted edges. Based on the graph and lifted edges, we formulate a lifted Multicut graph partitioning problem and solve it with the hierarchical Lifted Multicut solver introduced in Pape et al. (*87*) as an extension of the Multicut solver (*91*).

Note that solving a Multicut problem based on the graph without lifted edges can also yield a segmentation. However, we observe that this segmentation heavily over-segments microvilli and flagella. Both structures have an elongated shape with a small diameter, which results in small supervoxels. This results in a non-robust estimate for the boundary evidence-based edge weights and causes the over-segmentation. Hence, we add the attractive lifted edges derived from more robust semantic predictions to alleviate the degree of over-segmentation. Also note that the segmentation obtained via connected components on the thresholded predictions on its own is not of sufficient quality, mainly because it falsely merges several cells. As the last step, we manually proofread objects involved in the interactions studied closer using Paintera (https://github.com/saalfeldlab/paintera), and in some cases manually merged the microvilli from a single choanocyte collar into a single object.

Segmentation and snapshots of the high-resolution neuroid cells were obtained using Amira (Thermo Fisher Scientific).

FIB-SEM data access via MoBIE

The FIB-SEM datasets presented in Figure 6 and Supplementary Figure 28 are accessible via the FIJI plugin MoBIE (*92*). Views for panels D-E of Figure 6, and panels D, G, H, and I of Supplementary Figure 28 can be retrieved using MoBIE's bookmark function. The data is available at https://github.com/mobie/sponge-fibsem-project and can be accessed by providing

this address to the MoBIE plugin. See https://mobie.github.io/ for details on how to install and use MoBIE.

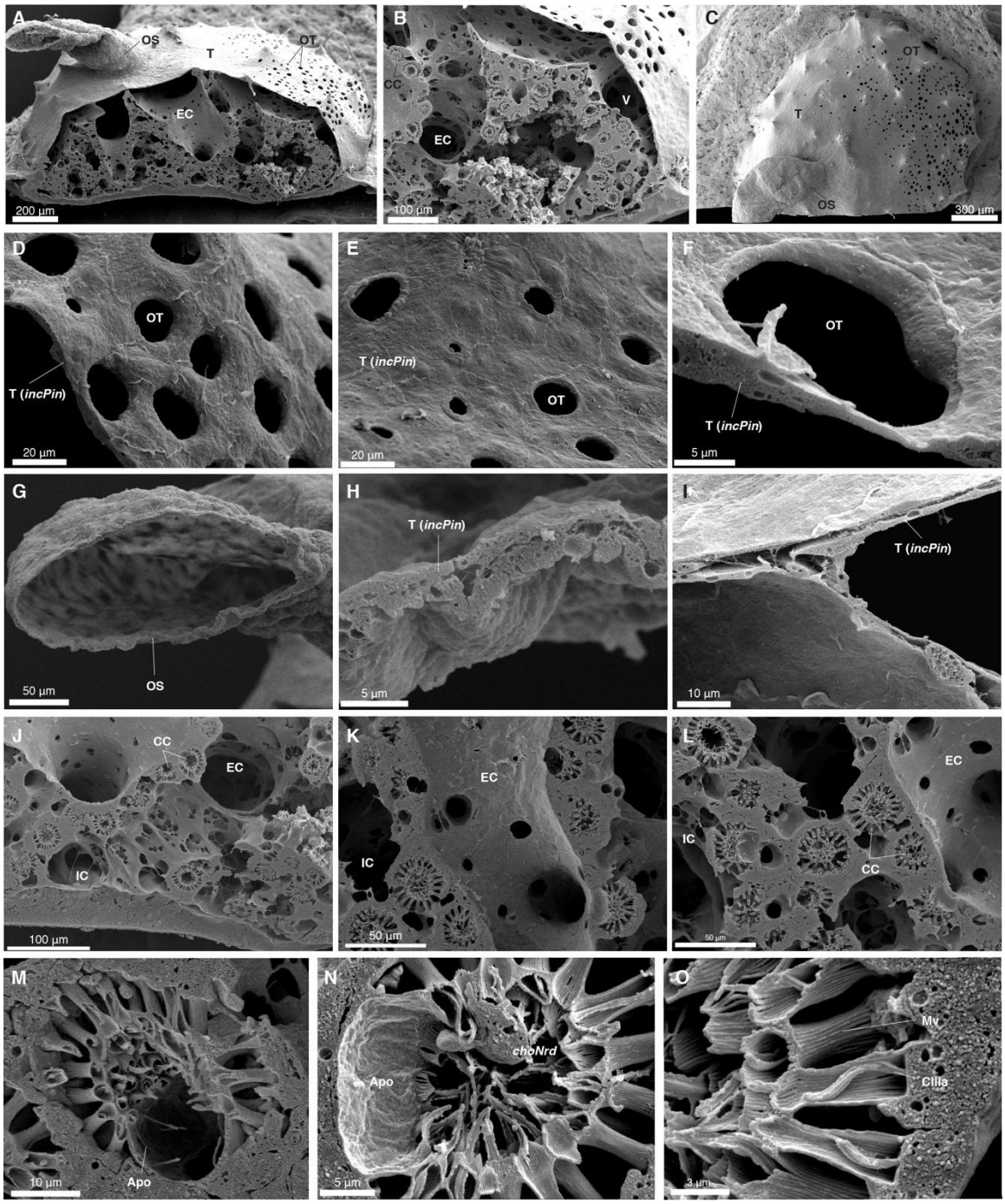

Figure S1. SEM reveals general morphology of *Spongilla lacustris*.

(A-B) Lateral view cross-section of juvenile *S. lacustris*. CC – choanocyte chamber, EC – excurrent canal, OS – osculum, OT – ostia, T – epithelial tent. V – vestibule. Water enters through ostia into the vestibule and is drawn through the canal system via choanocyte chambers before exiting through the osculum. **(C)** Overhead view. **(D-F)** Images of ostia, incurrent openings in the outer epithelial tent through which water enters the vestibule. **(G)** Images of the osculum, the final excurrent structure through which water exits the sponge. **(H-I)** Cross-sections of the outer epithelial tent. **(J-L)** Choanocyte chambers embedded in the canal system. IC –

incurrent canal. **(M-N)** Choanocyte chambers with ciliated apopylar cells forming an excurrent pore. Apo – apopylar pore. **(O)** Cross-section of choanocyte microvilli collars and cilia. Mv – microvillus.

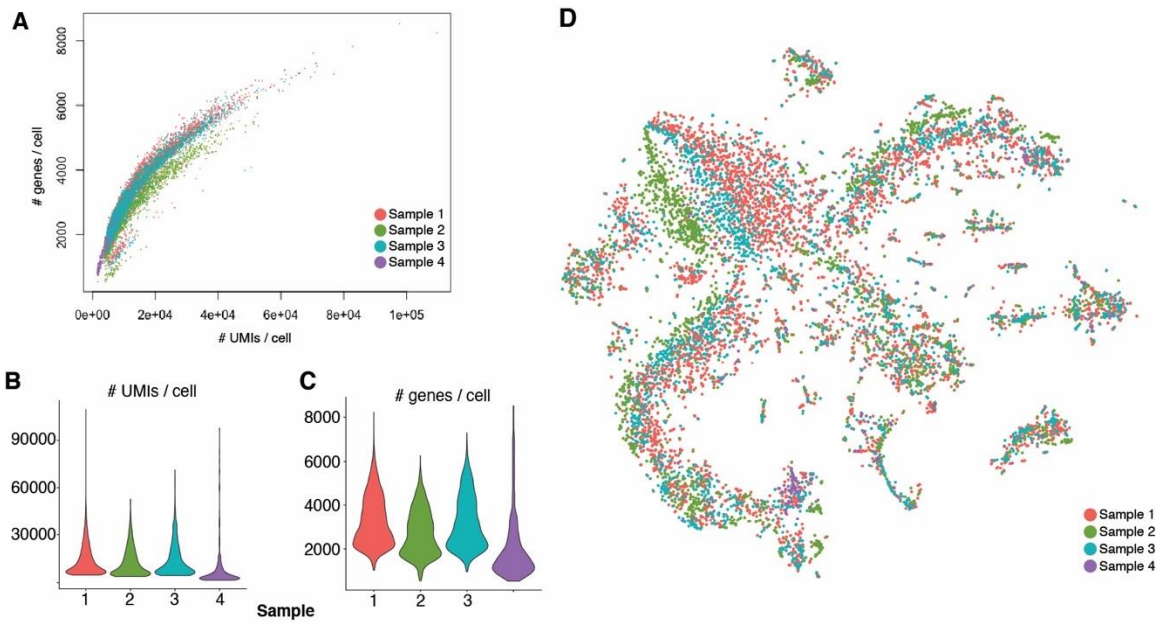

Figure S2. Sample statistics from whole-body scRNAseq of *S. lacustris*.

(A) UMI count vs. expressed gene count for all cells (colored by sample). **(B)** Violin plots of UMI count distribution for each sample. **(C)** Violin plots of gene count distribution for each sample. **(D)** tSNE plot with cells colored by sample.

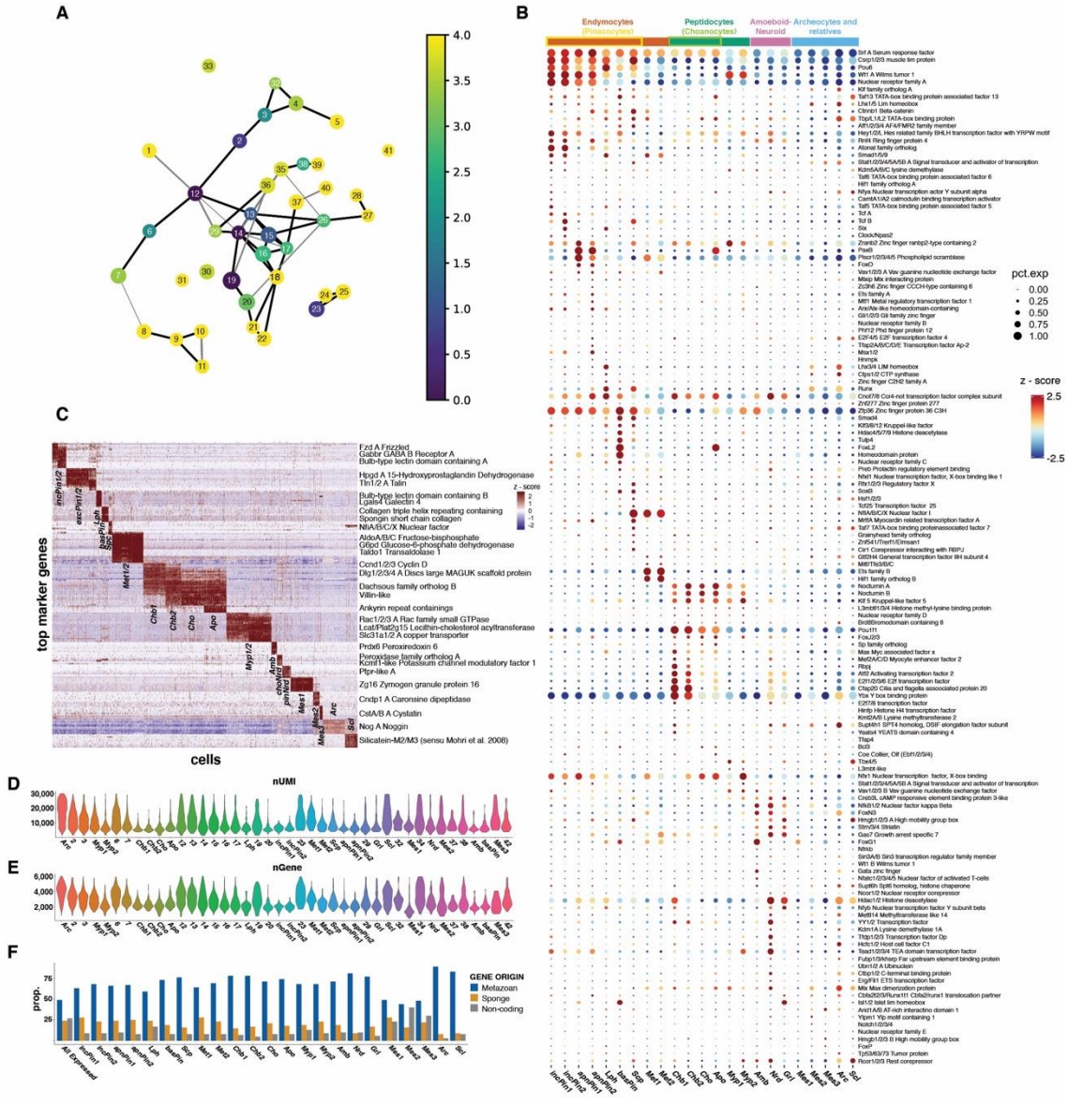

Figure S3. Clustering of whole-body scRNAseq of *S. lacustris*.

(A) PAGA connectivity plot with clusters colored by the log number of specific marker genes. Clusters with more than 30 specific markers (> 3.4 log value) represent distinct expression programs and were tied to morphological cell types. **(B)** Dotplot showing all expressed transcription factors in *Spongilla*. **(C)** Heatmap of top markers for *Spongilla* cell types. **(D)** Violin plots of UMI (transcript) count distribution for each of 42 distinct clusters. **(E)** Violin plots of gene count distribution for each of 42 distinct clusters. **(F)** Gene classes for top 100 markers from each cell type. Metazoan – protein-coding gene originating in an animal ancestor. Sponge – protein-coding genes originating within Porifera or in poriferan stem line. Non-coding – does not contain at least 70 amino acid open reading frames.

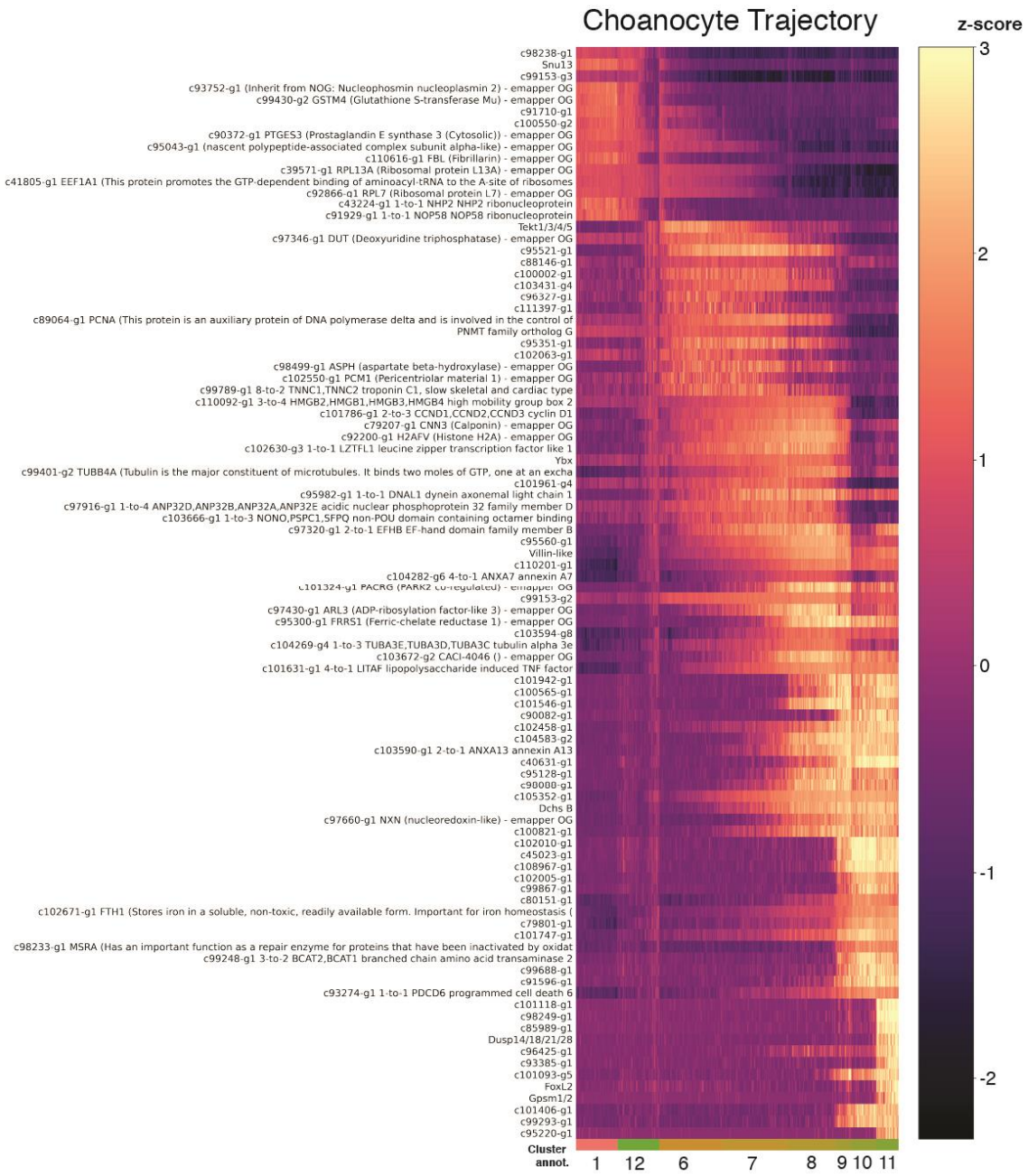

Figure S4. Choanocyte lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to choanocytes. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

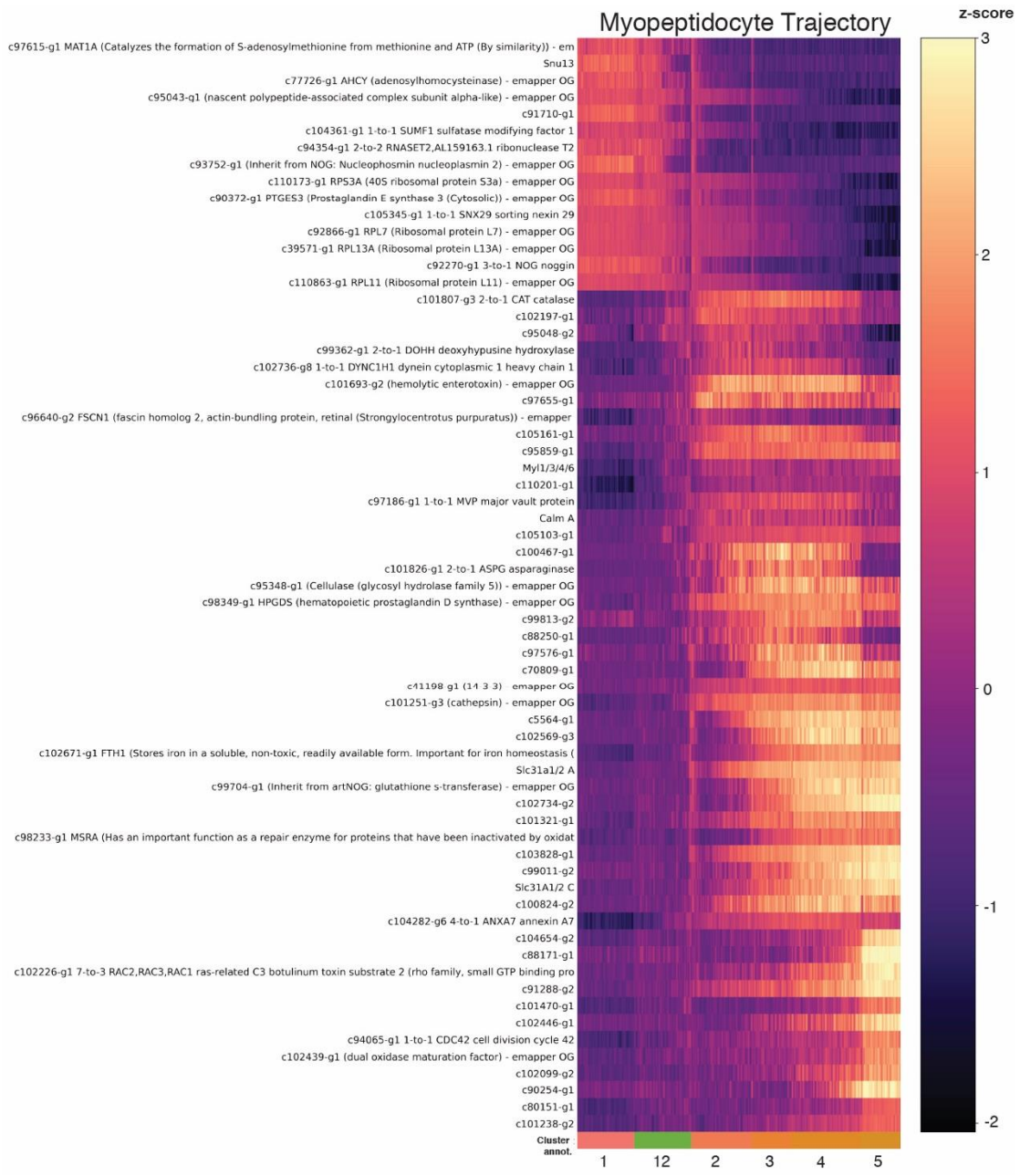

Figure S5. Myopeptidocyte lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to myopeptidocytes. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

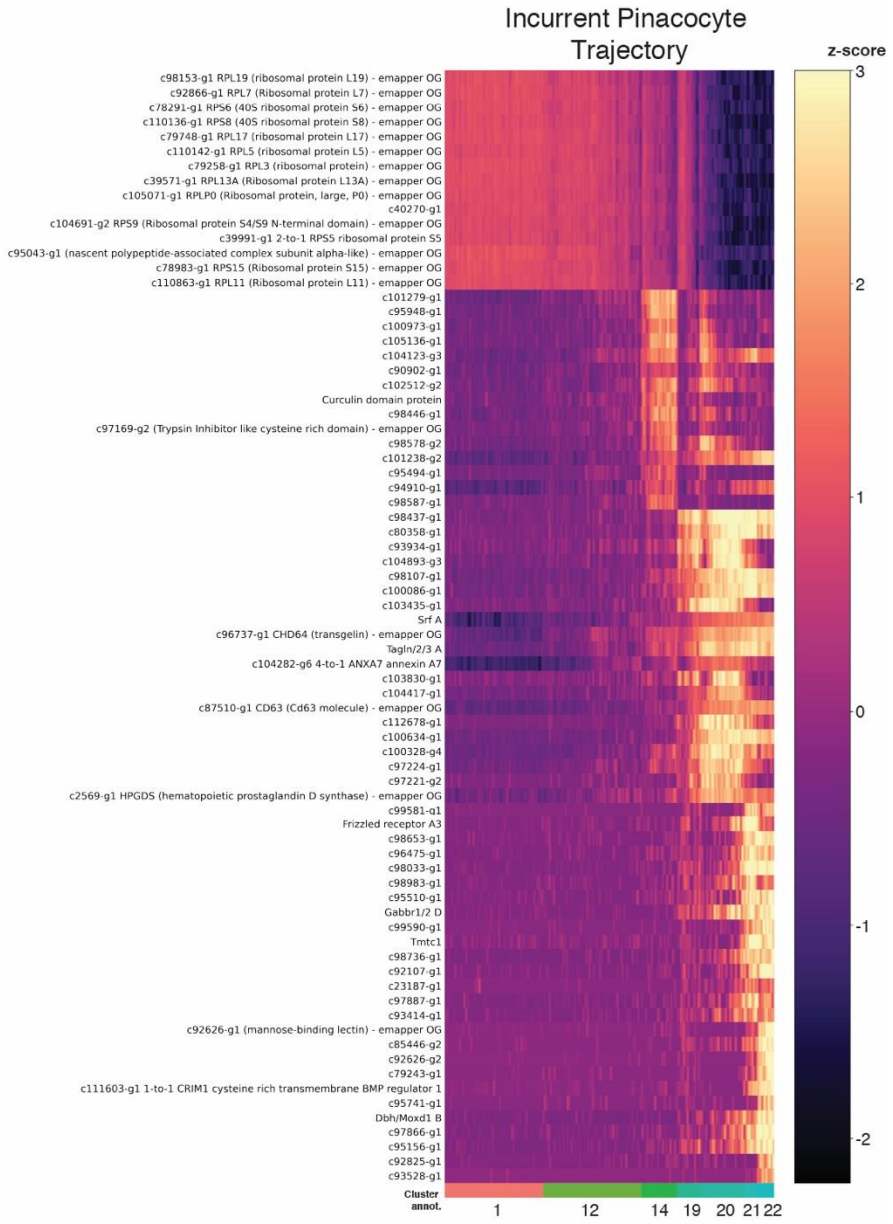

Figure S6. Incurrent pinacocyte lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to pinacocytes. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

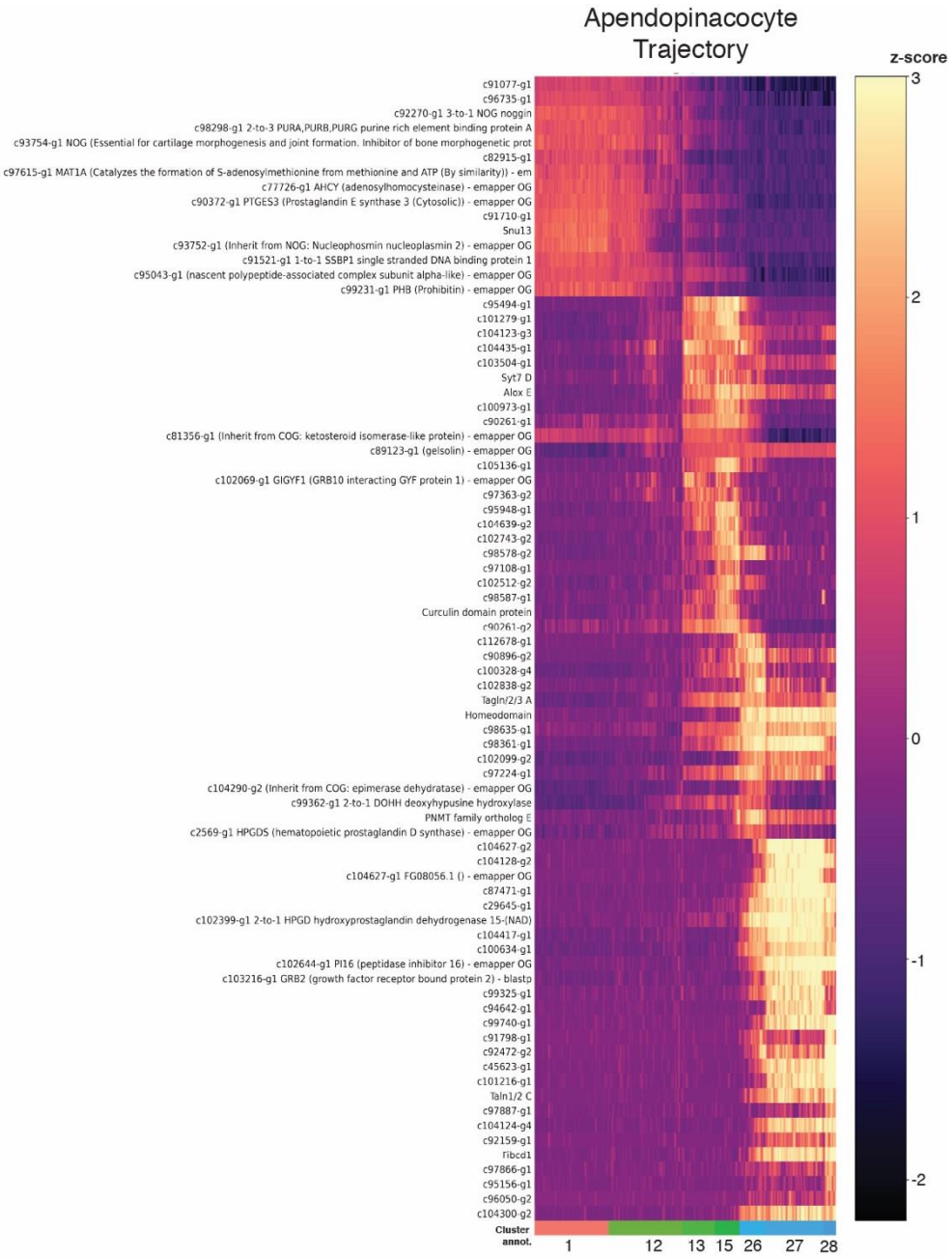

Figure S7. Apendopinacocyte lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to apendopinacocytes. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

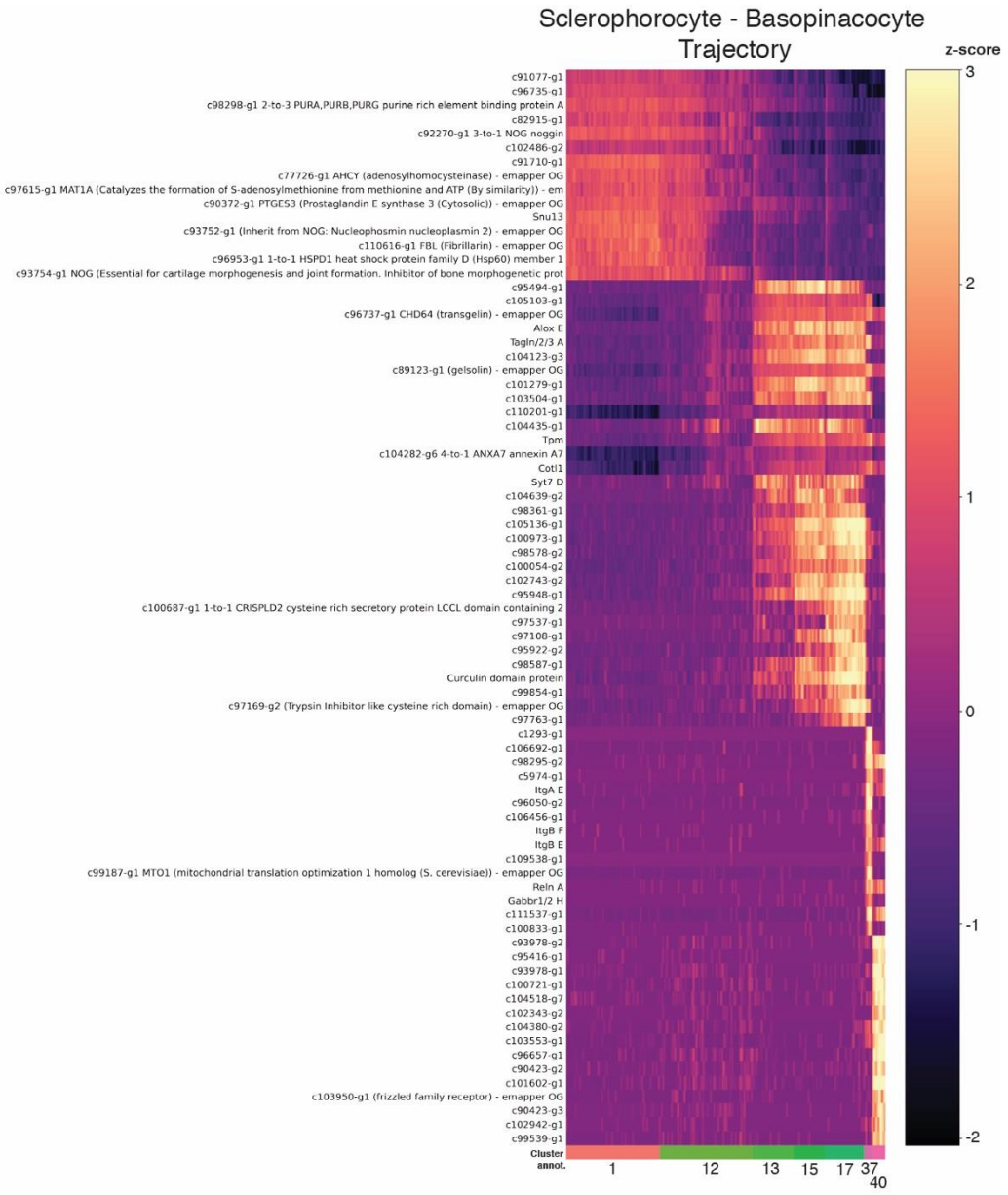

Figure S8. Sclerophorocyte-Basopinacocyte lineage top differentially expressed genes.
Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to sclerophorocytes and basopinacocytes. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

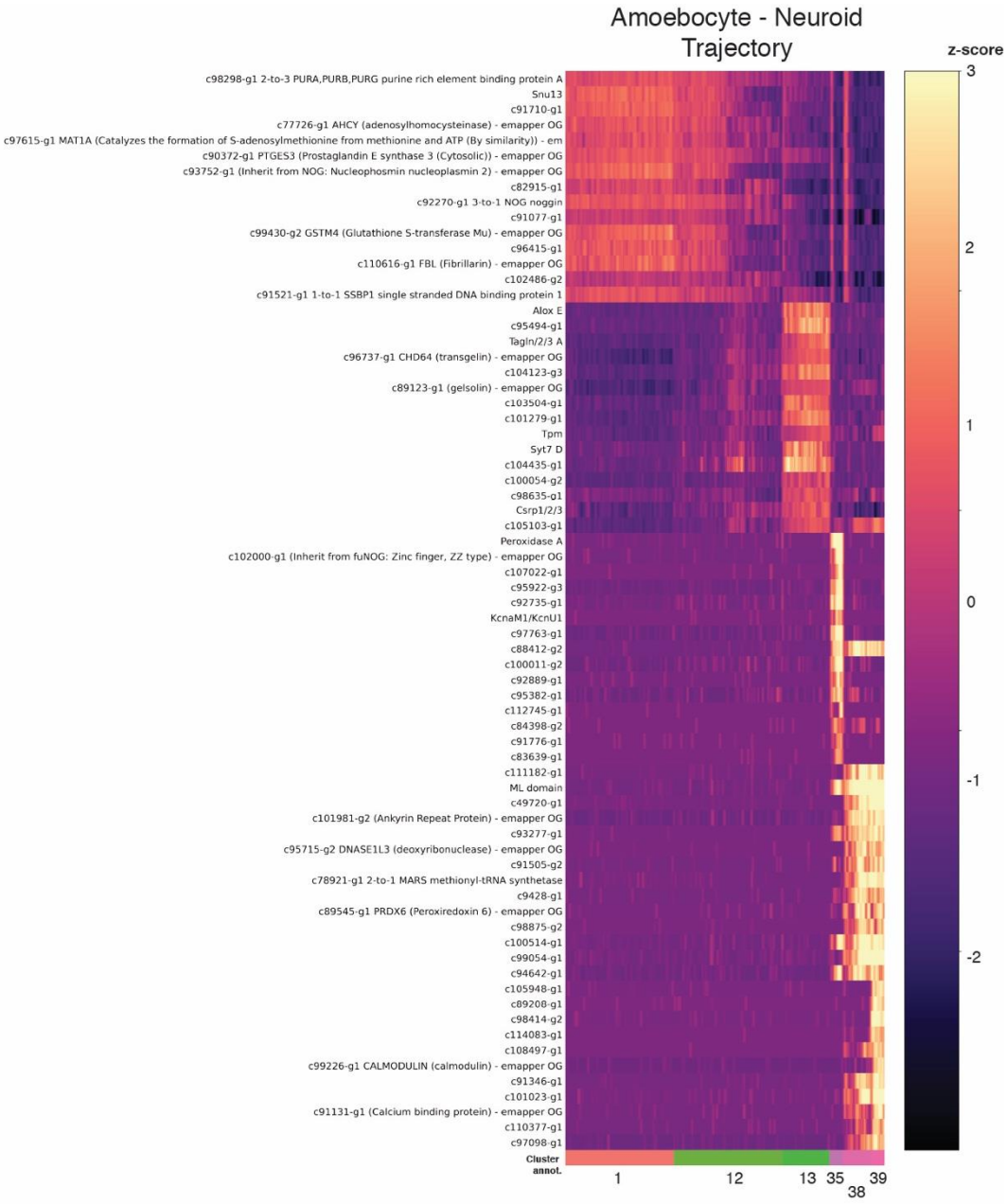

Figure S9. Amoeboecyte-Neuroid lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to amoebocytes and neuroid cells. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

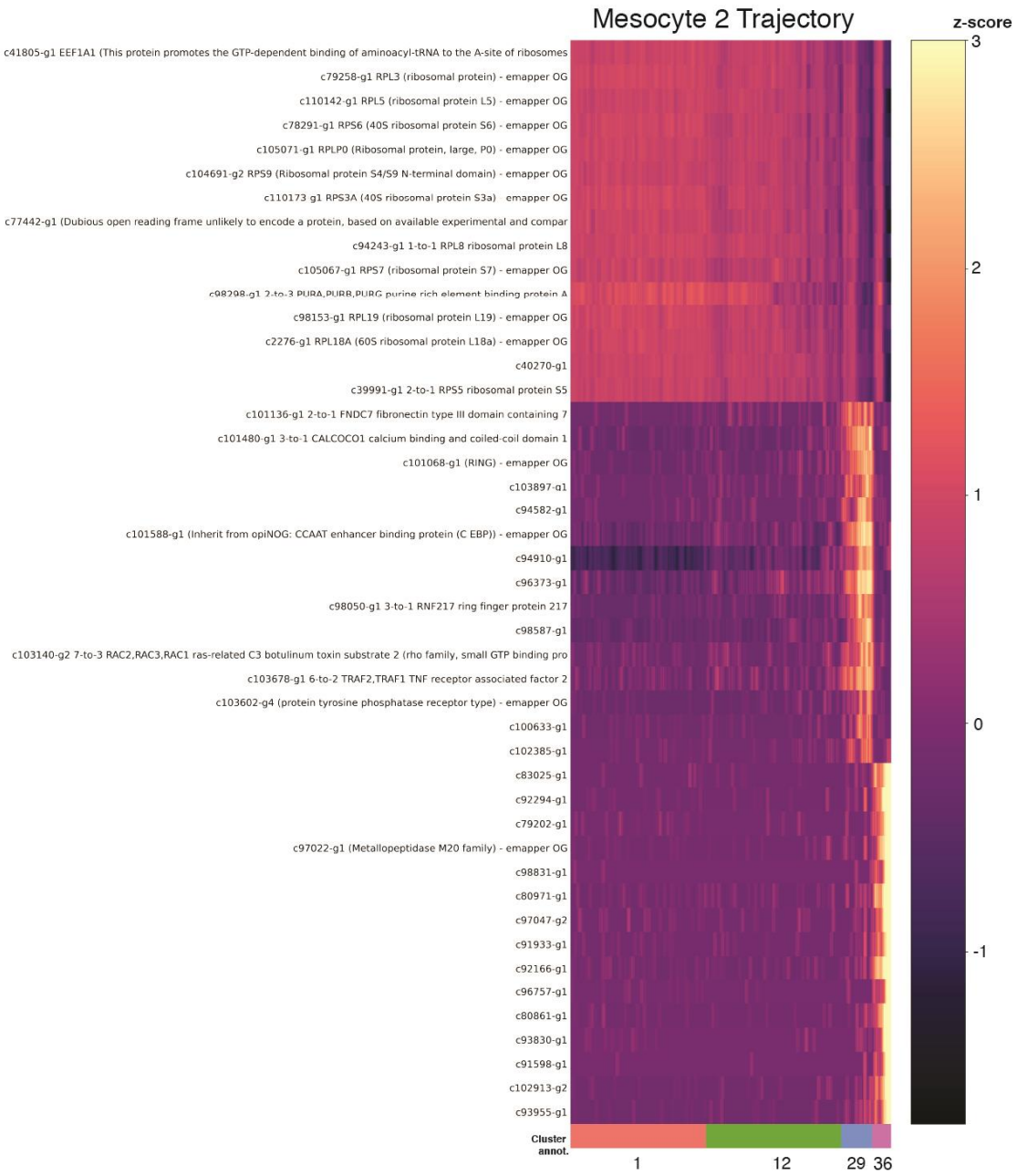

Figure S10. Mesocyte 2 lineage top differentially expressed genes.

Heatmap showing top differentially expressed genes for PAGA-based developmental trajectory from archaeocytes to mesocytes 2. Genes beginning with a gene ID (e.g. c###-g#) represent automated names (see methods). Genes lacking a gene ID and only containing a gene name have been manually-curated. Gene's with only a gene ID and no following information are either protein-coding genes that lack homology information (e.g. sponge-specific genes), or are non-coding.

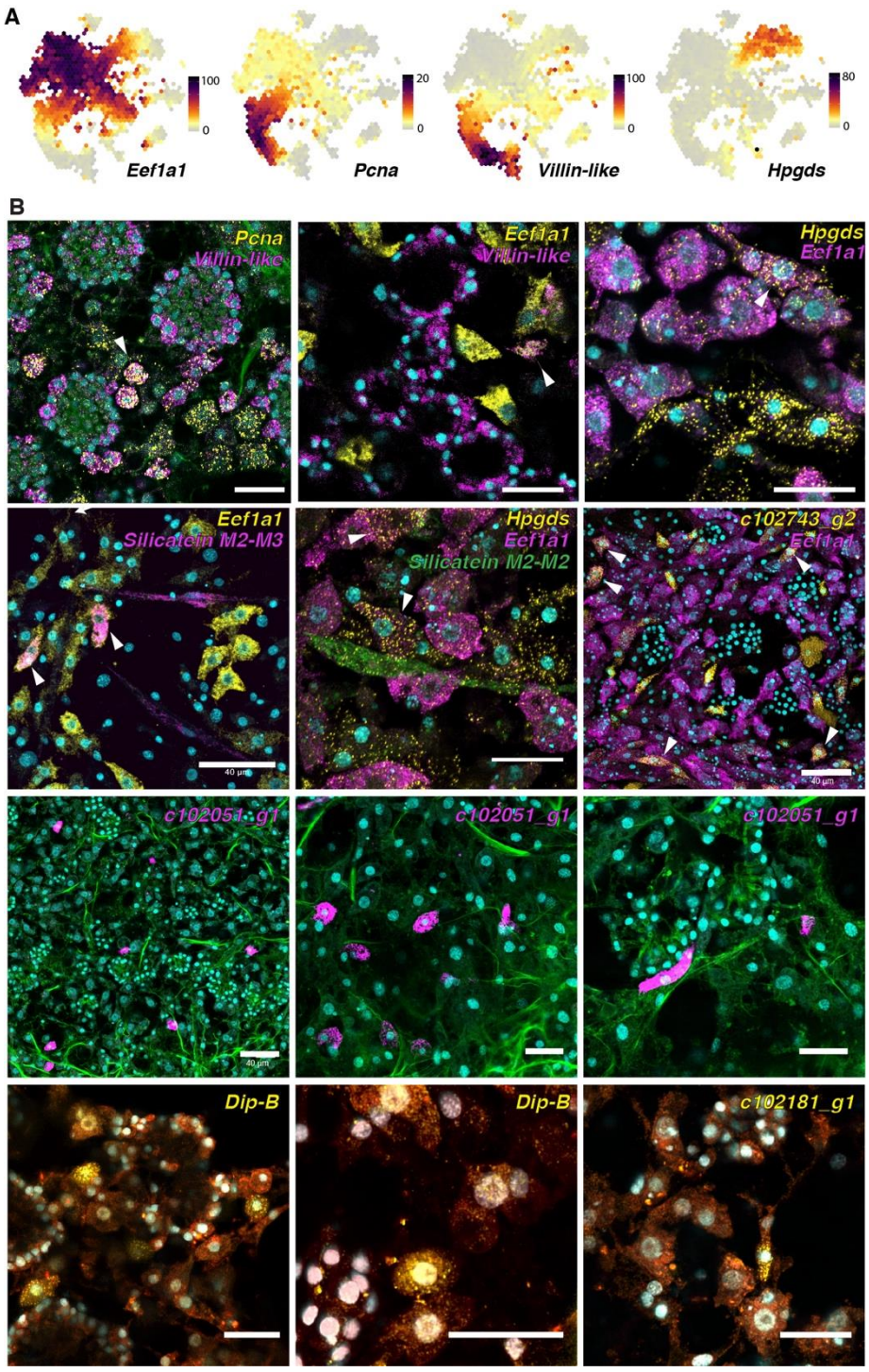

Figure S11. smFISH of transitional clusters and cell types in the archaeocyte and relatives family.

(A) Hexplots showing average expression (counts per ten thousand) of archaeocyte (*Eef1a1*), choanoblast (*Pcna*), choanocyte (*Villin-like*), and myopeptidocyte (*Hpgds*) marker genes. **(B)** smFISH of archaeocyte, myopeptidocyte, choanocyte, choanoblast, and mesocyte markers. Transitional progenitor cells (arrow) co-express archaeocyte and differentiated cell type markers.

co-expressing archaeocyte and differentiated showing differentiated and transitional (arrow) cells with observed co-localization of selected gene markers. *Eef1a1* labels archaeocytes with prominent nucleolus as well as putative transitional cells co-stained with markers for other cell types, including sclerocytes (*Silicatein M2-M3*), myopeptidocytes (*Hpgds*), and pinacocyte progenitors (*c102743_g2*). Mesocyte 1 marker (c102051_g1) stains medium-to-large cells scattered throughout sponge mesohyl. Mesocyte 2 marker (ortholog of Drosophila *Dip-B*) identifies medium-sized ovoid-shaped cells in the mesohyl. Mesocyte 3 marker (*c102181_g1*) stains very rare small cells. Nuclei stained with DAPI (cyan). Membrane stains are CellBrite Fix (green) or Fm143-Fx (red). CellBrite Fix also labels intercellular fibrils. Scale bar 20 μm unless otherwise noted.

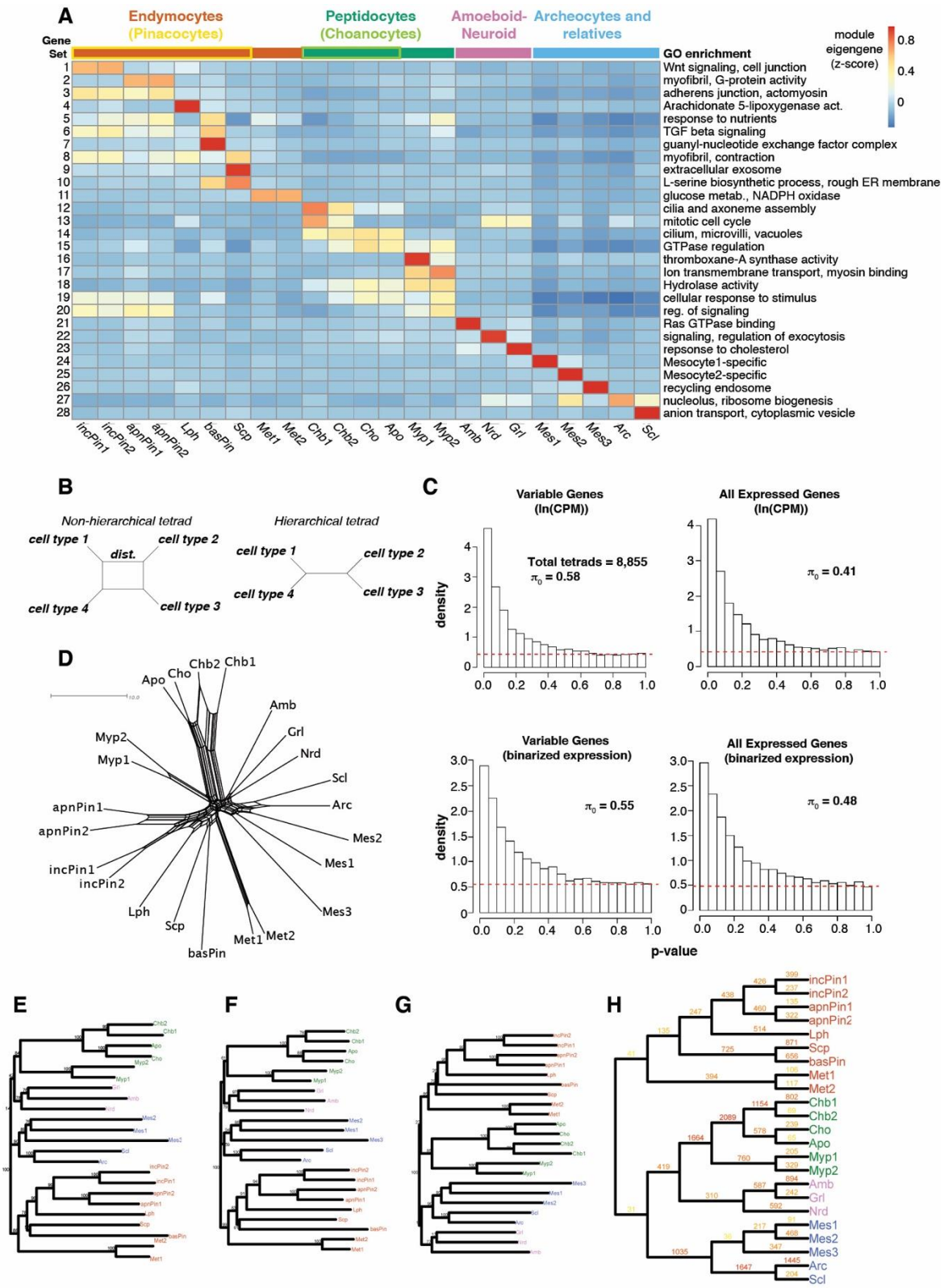

Figure S12. Hierarchical structure of whole-body scRNAseq in *S. lacustris*.
(A) Heatmap showing eigengene expression for all WGCNA gene sets. **(B)** Box plots depicting examples of non-hierarchical (top) and hierarchical (bottom) relationships of cell type tetrads.

Tetrad graphs are computed from the distances between each of the four cell type's average expression profile. **(C)** P-value distributions from treeness tests of 8,855 cell type tetrads. Red dashed line indicates uniform distribution of p-values expected under null hypothesis. **(D)** SplitsTree network of *S. lacustris* cell type relationships. **(E-G)** Neighbour-joining trees based on variable gene sets representing (E) the 5945 most variable genes (representing the *Spongilla* cell type tree in Fig. 2A), (F) the 2454 most variable genes, and (G) the 11,178 most variable genes. **(H)** *Spongilla* cell type cladogram showing the number of genes inferred to change optimal expression levels in each clade.

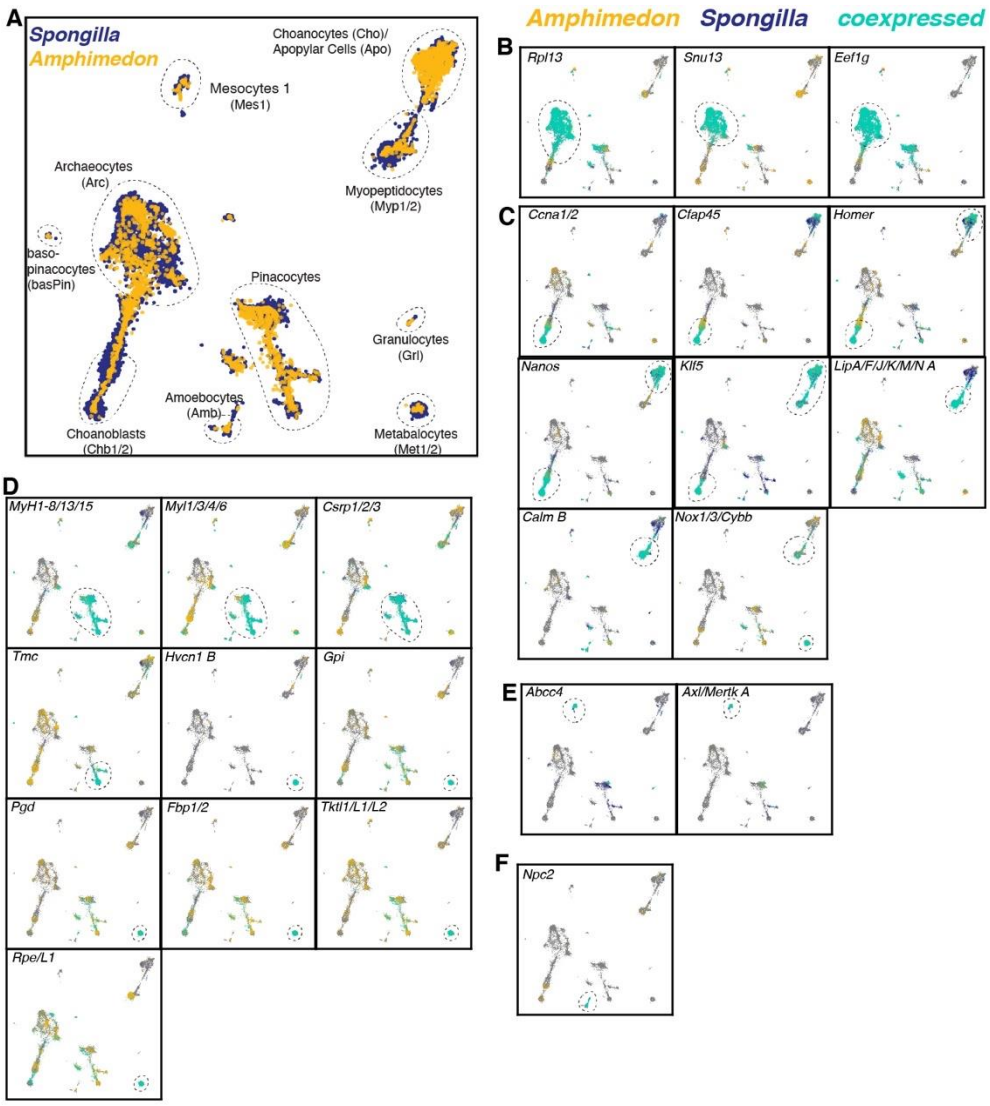

Figure S13. *Spongilla* and adult *Amphimedon* conserved cell type markers.

(A) UMAP projection of the combined *Spongilla* (blue) and adult *Amphimedon* (yellow) manifolds. Select cell types and cell type clades circled. **(B-F)** Expression of SAMap-linked orthologous gene pairs, overlaid on the combined species UMAP projection. Cells expressing ortholog are color-coded by species, with cells expressing orthologs and connected across species colored in cyan. Cells lacking expression are colored gray. (B) Archaeocyte marker genes. (C) Peptidocyte marker genes. (D) Endymocyte marker genes. (E) Mesocyte 1 marker genes. (F) Granulocyte marker genes.

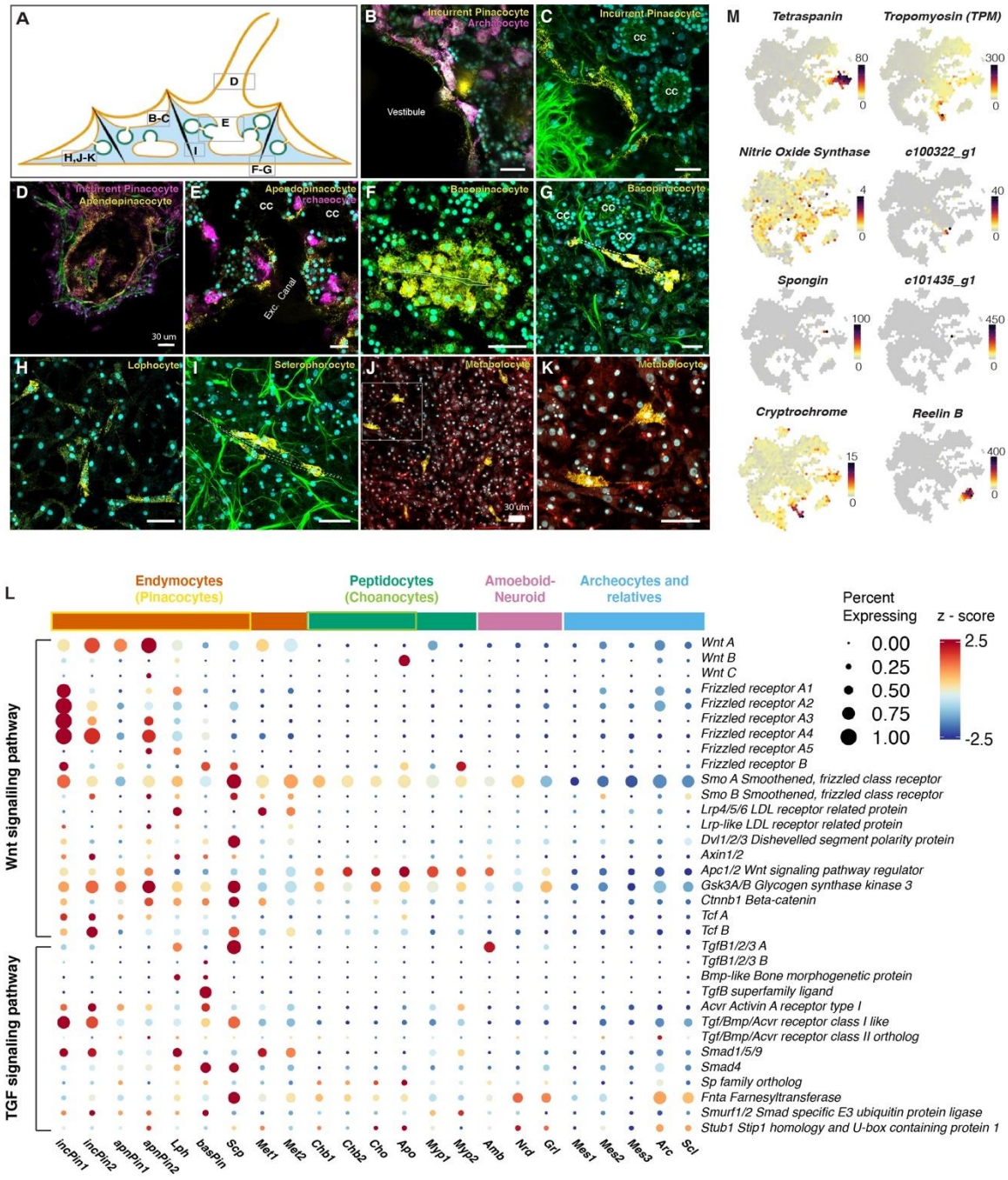

Figure S14. Endymocyte family.

(A) Diagram of juvenile sponge illustrating the location of smFISH panels. **(B-K)** smFISH of endymocyte cell type markers. Markers: (B) *Tropomyosin* and *Eef1a1*, (C) *Fibcd1 A*, (D) *Tropomyosin* and *c99740_g1*, (E) *Tetraspanin* and *Eef1a1*, (F-G) *Spongin short-chain collagen*, (H) *c100322_g1*, (I) *c101435_g1*, (J-K) *Reelin B*. All scale bars 20μm except where noted in the panel. Membrane stains are CellBrite Fix (green) or Fm143-Fx (red); nuclei DAPI stain (cyan). Dashed lines show the outline of the spicule. CC – choanocyte chamber. **(L)** Dotplot of all

expressed Wnt and Tgf signaling pathway genes. **(M)** Hexplots showing average expression of endymocyte smFISH marker genes in Fig. 3D.

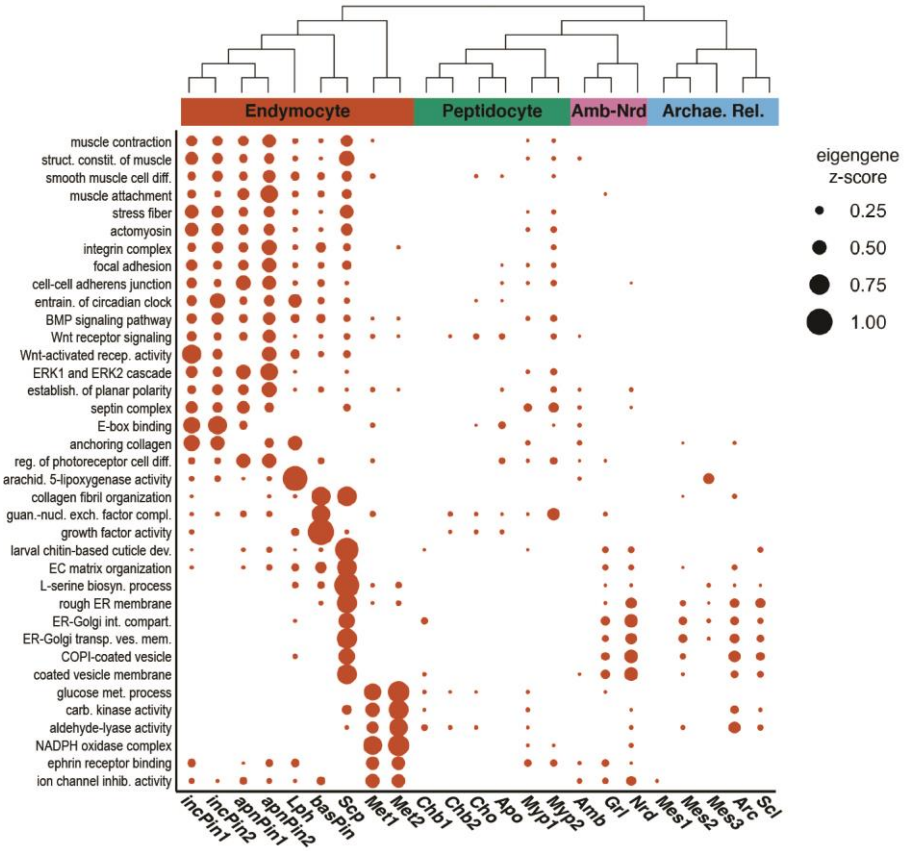

Figure S15. Endymocyte family GO term dotplot.

Dotplot of select GO terms enriched in endymocyte cell types and clades.

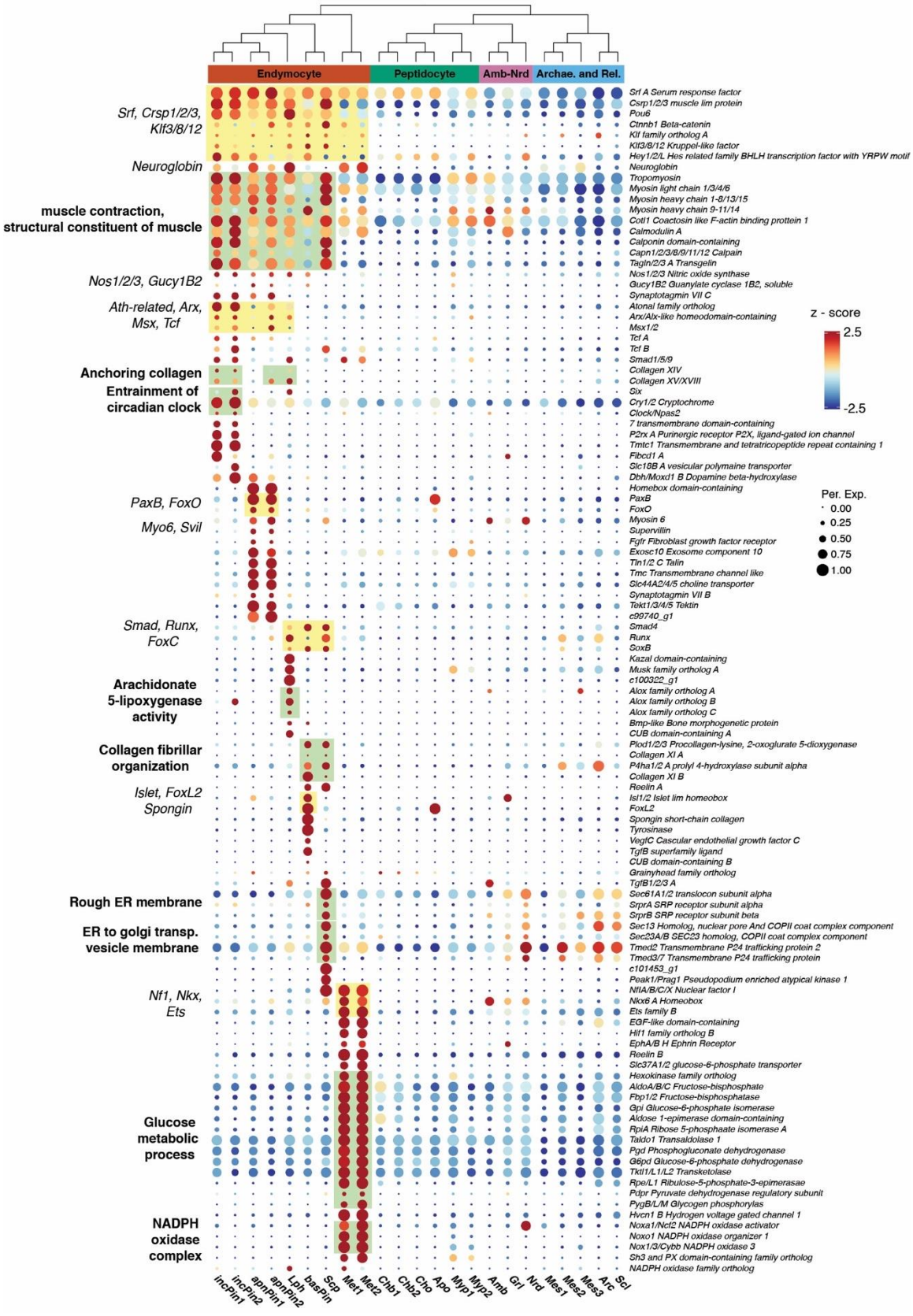
Endymocyte
Peptidocyte
Amb-Nrd
Archae. and Rel.
Srf, Crsp1/2/3, Klf3/8/12
Neuroglobin
muscle contraction, structural constituent of muscle
Nos1/2/3, Gucy1B2
Ath-related, Arx, Msx, Tcf
Anchoring collagen
Entrainment of circadian clock
PaxB, FoxO
Myo6, Svil
Smad, Runx, FoxC
Arachidonate 5-lipoxygenase activity
Collagen fibrillar organization
Islet, FoxL2 Spongin
Rough ER membrane
ER to golgi transp. vesicle membrane
Nf1, Nkx, Ets
Glucose metabolic process
NADPH oxidase complex
Srf A Serum response factor
Csrp1/2/3 muscle lim protein
Pou6
Ctnnb1 Beta-catenin
Klf family ortholog A
Klf3/8/12 Kruppel-like factor
Hey1/2/L Hes related family BHLH transcription factor with YRPW motif
Neuroglobin
Tropomyosin
Myosin light chain 1/3/4/6
Myosin heavy chain 1-8/13/15
Myosin heavy chain 9-11/14
Cotl1 Coactosin like F-actin binding prottein 1
Calmodulin A
Calponin domain-containing
Capn1/2/3/8/9/11/12 Calpain
Tagln/2/3 A Transgelin
Nos1/2/3 Nitric oxide synthase
Gucy1B2 Guanylate cyclase 1B2, soluble
Synaptotagmin VII C
Atonal family ortholog
Arx/Alx-like homeodomain-containing
Msx1/2
Tcf A
Tcf B
Smad1/5/9
Collagen XIV
Collagen XV/XVIII
Six
Cry1/2 Cryptochrome
Clock/Npas2
7 transmembrane domain-containing
P2rx A Purinergic receptor P2X, ligand-gated ion channel
Tmtc1 Transmembrane and tetratricopeptide repeat containing 1
Fibcd1 A
Slc18B A vesicular polymaine transporter
Dbh/Moxd1 B Dopamine beta-hydroxylase
Homebox domain-containing
PaxB
FoxO
Myosin 6
Supervillin
Fgfr Fibroblast growth factor receptor
Exosc10 Exosome component 10
Tln1/2 C Talin
Tmc Transmembrane channel like
Slc44A2/4/5 choline transporter
Synaptotagmin VII B
Tekt1/3/4/5 Tektin
c99740_g1
Smad4
Runx
SoxB
Kazal domain-containing
Musk family ortholog A
c100322_g1
Alox family ortholog A
Alox family ortholog B
Alox family ortholog C
Bmp-like Bone morphogenetic protein
CUB domain-containing A
Plod1/2/3 Procollagen-lysine, 2-oxoglurate 5-dioxygenase
Collagen XI A
P4ha1/2 A prolyl 4-hydroxylase subunit alpha
Collagen XI B
Reelin A
Isl1/2 Islet lim homeobox
FoxL2
Spongin short-chain collagen
Tyrosinase
VegfC Cascular endothelial growth factor C
TgfB superfamily ligand
CUB domain-containing B
Grainyhead family ortholog
TgfB1/2/3 A
Sec61A1/2 translocon subunit alpha
SrprA SRP receptor subunit alpha
SrprB SRP receptor subunit beta
Sec13 Homolog, nuclear pore And COPII coat complex component
Sec23A/B SEC23 homolog, COPII coat complex component
Tmed2 Transmembrane P24 trafficking protein 2
Tmed3/7 Transmembrane P24 trafficking protein
c101453_g1
Peak1/Prag1 Pseudopodium enriched atypical kinase 1
NfIA/B/C/X Nuclear factor I
Nkx6 A Homeobox
Ets family B
EGF-like domain-containing
Hif1 family ortholog B
EphA/B H Ephrin Receptor
Reelin B
Slc37A1/2 glucose-6-phosphate transporter
Hexokinase family ortholog
AldoA/B/C Fructose-bisphosphate
Fbp1/2 Fructose-bisphosphatase
Gpi Glucose-6-phosphate isomerase
Aldose 1-epimerase domain-containing
RpiA Ribose 5-phosphaate isomerase A
Taldo1 Transaldolase 1
Pgd Phosphogluconate dehydrogenase
G6pd Glucose-6-phosphate dehydrogenase
Tktl1/L1/L2 Transketolase
Rpe/L1 Ribulose-5-phosphate-3-epimerasae
Pdpr Pyruvate dehydrogenase regulatory subunit
PygB/L/M Glycogen phosphorylas
Hvcn1 B Hydrogen voltage gated channel 1
Noxa1/Ncf2 NADPH oxidase activator
Noxo1 NADPH oxidase organizer 1
Nox1/3/Cybb NADPH oxidase 3
Sh3 and PX domain-containing family ortholog
NADPH oxidase family ortholog
z - score
2.5
-2.5
Per. Exp.
0.00
0.25
0.50
0.75
1.00
incPin1
incPin2
apnPin1
apnPin2
Lph
basPin
Scp
Met1
Met2
Chb1
Chb2
Cho
Apo
Myp1
Myp2
Amb
Gri
Nrd
Mes1
Mes2
Mes3
Arc
Scl

Figure S16. Endymocyte family enriched genes.

Boxes around dots delineate expression modules of transcription factors (yellow) or functional complexes/pathways (green). Notable cell type or cell type clade differentially-expressed genes are shown in italics, whereas enriched GO-terms and a selection of differentially-expressed genes driving this enrichment, are labelled in bold.

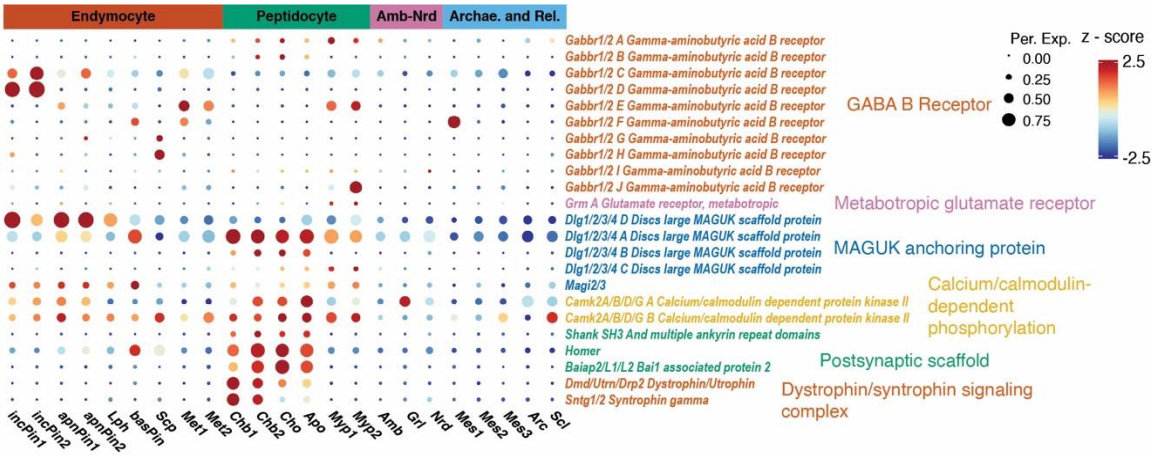

Figure S17. Expression of 'post'-synaptic/receptive homologs in *Spongilla* cell types.

Dotplots showing scaled expression for orthologs of 'post'-synaptic/receptive genes in *Spongilla*. Diameter of circle indicates percent of cells within a cell type that are expressing the gene.

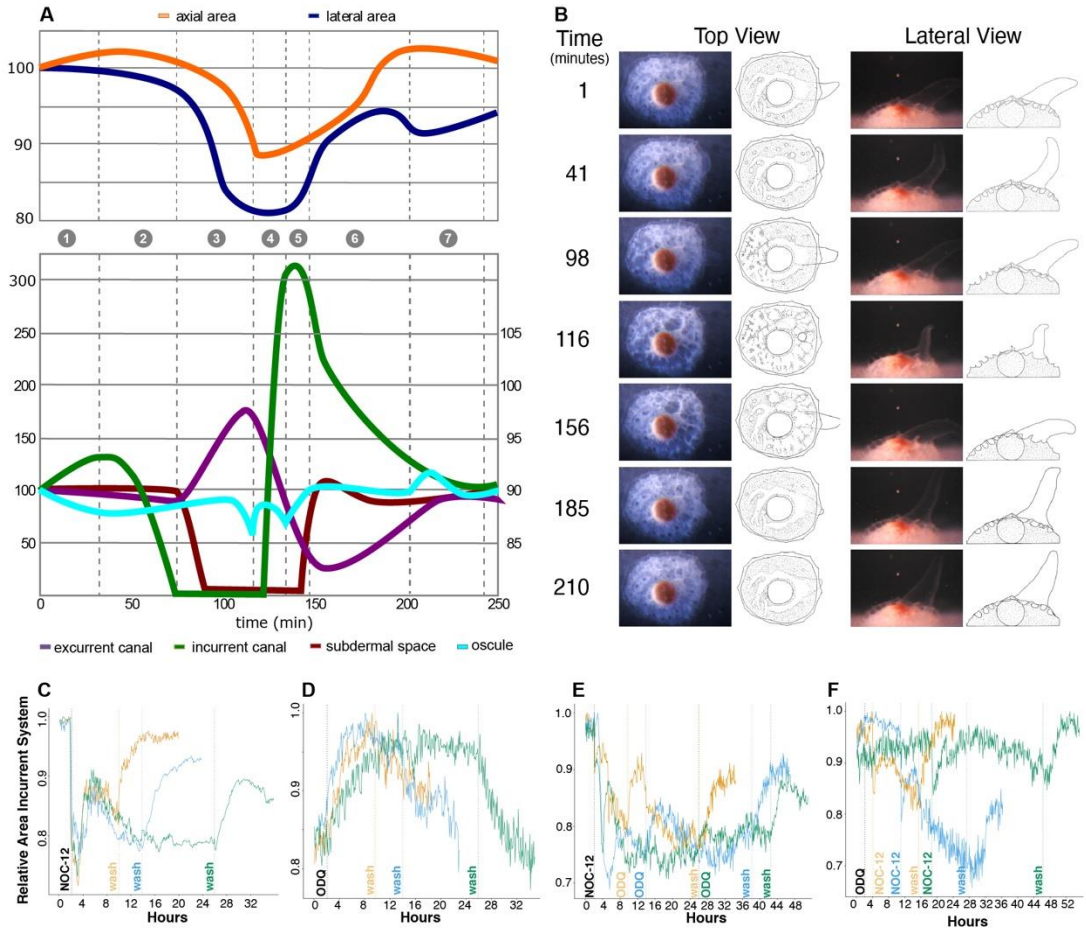

Figure S18. Contractions in juvenile *S. lacustris*.

(A) Graph showing changes in the relative area of canal system components during 7 stages of a typical contraction cycle. Colored lines depict relative area of different regions of the canal system. **(B)** Brightfield and schematic views of sponge for each of 7 stages of typical contraction cycle. Times (in minutes) correspond to those in the graph in panel A.

(C-F) Timelapse plot of contraction after addition of NOC-12 (C), ODQ (D), or both treatments in series (E-F). Colored lines indicate experiments with different periods between successive treatments (see methods).

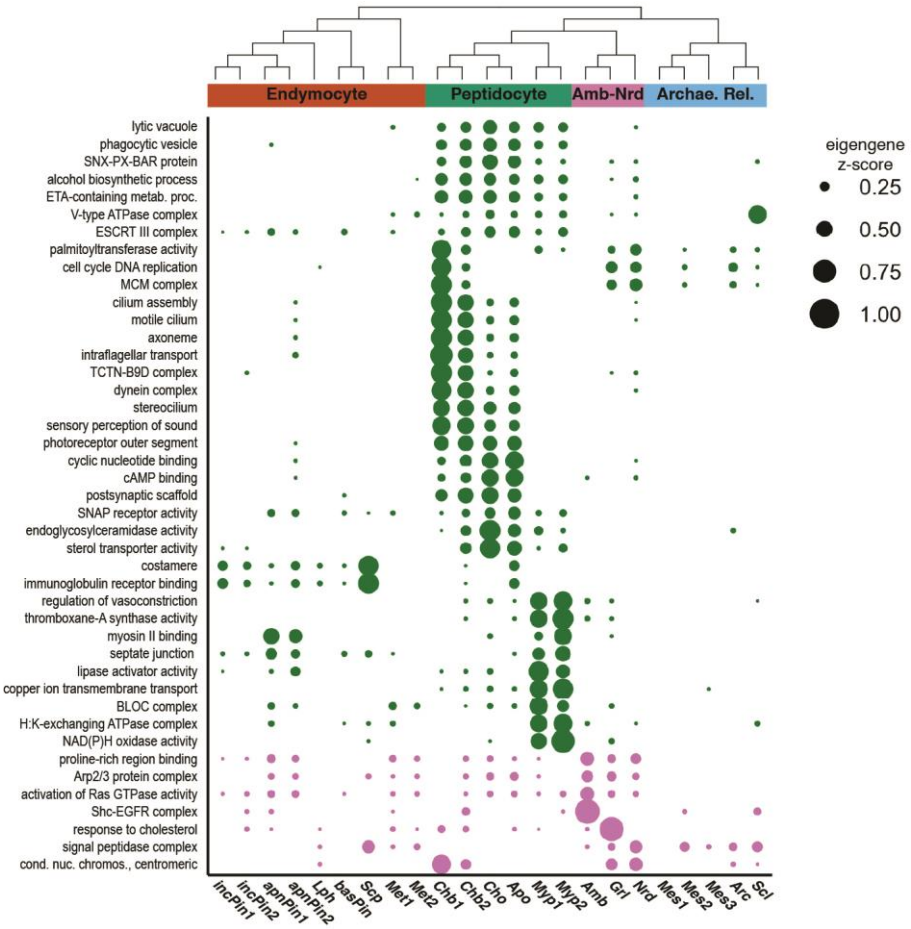

Figure S19. Peptidocyte and Neuroid-Amoeboid families GO term dotplot.

Dotplot of select GO terms enriched in Peptidocyte and Neuroid-Amoeboid cell types and clades.

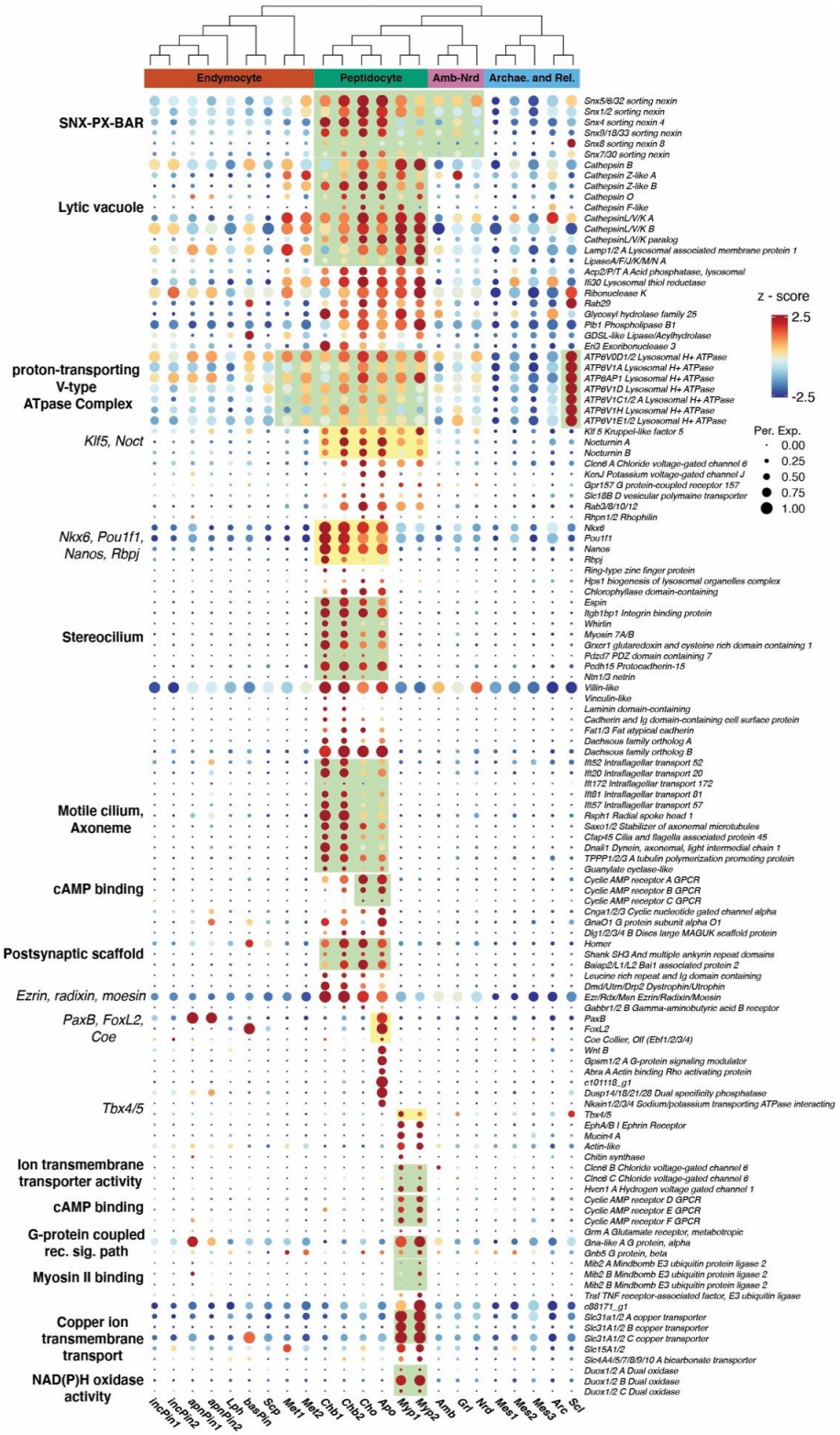

Figure S20. Peptidocyte family-enriched genes.

Boxes around dots delineate expression modules of transcription factors (yellow) or functional complexes/pathways (green). Notable cell type or cell type clade differentially-expressed genes

are shown in italics, whereas enriched GO-terms and a selection of differentially-expressed genes driving this enrichment, are labelled in bold.

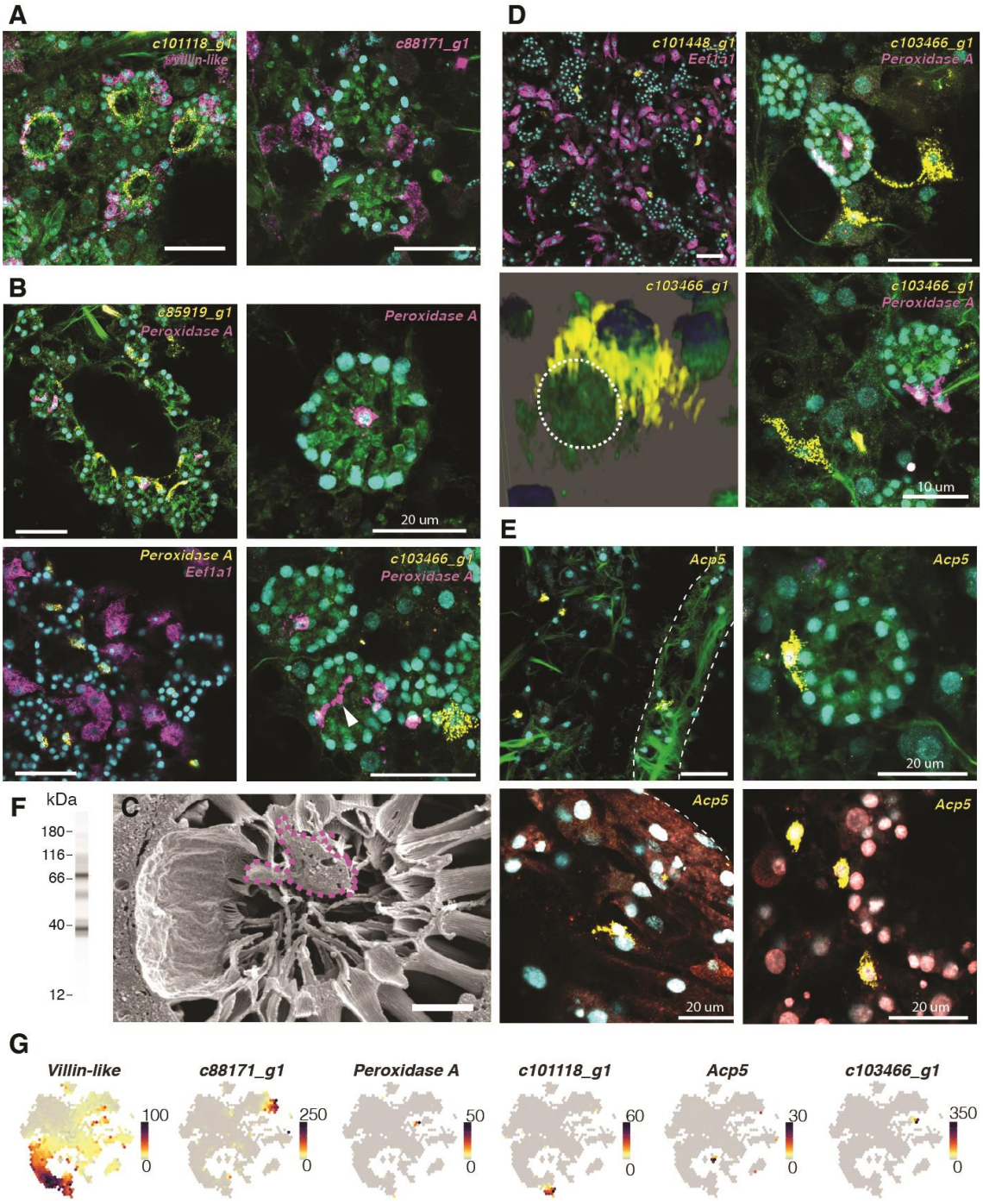

Figure S21. Peptidocyte and Neuroid-amoeboid Family smFISH.

(A) smFISH of choanocytes (*Villin-like*), apopylar cells (*c101118_g1*), and myopeptidocytes (*c88171_g1*). Membrane stains are CellBrite Fix (green) or Fm143-Fx (red); nuclei DAPI stain (cyan); Scale bar 30 μm unless otherwise noted. **(B)** smFISH of neuroid cells (*Peroxidase A*) with archaeocytes (*Eef1a1*) or apopylar cells (*c85989_g1*). Arrowhead denotes long filopodia-like extension from neuroid cell. **(C)** SEM of a neuroid cell (outlined by violet dashed line) in choanocyte chamber. **(D)** smFISH of amoebocytes (*c101448_g1* and *c103466_g1*) with

archaeocytes (*Eef1a1*) or neuroid cells (*Perodixase A*). Dotted white line outlines cell likely being engulfed by amoebocyte **(E)** smFISH of granulocytes (*Acp5*). **(F)** Western blot of anti-ERM antibody. **(G)** Hexplots showing average expression (counts per ten thousand) of peptidocyte and neuroid-amoeboid smFISH markers in Fig. 4A-F.

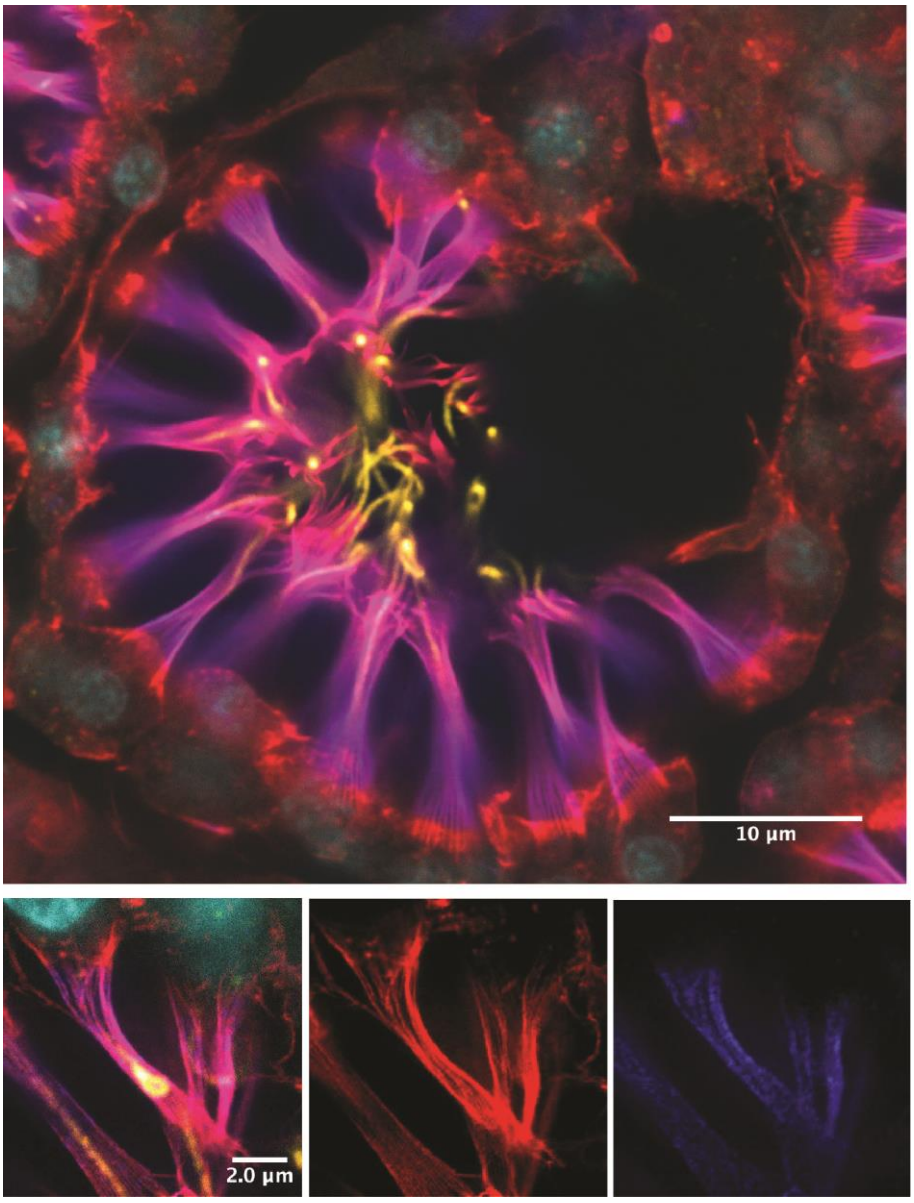

Figure S22. ERM protein in choanocyte chambers.

STED image of ezrin/radixin/moesin (ERM) protein localization in choanocyte chamber. Cyan – nuclei (DAPI), yellow – acetylated tubulin, red – F-actin (Phalloidin), blue – ERM protein, violet – overlap of F-actin and ERM.

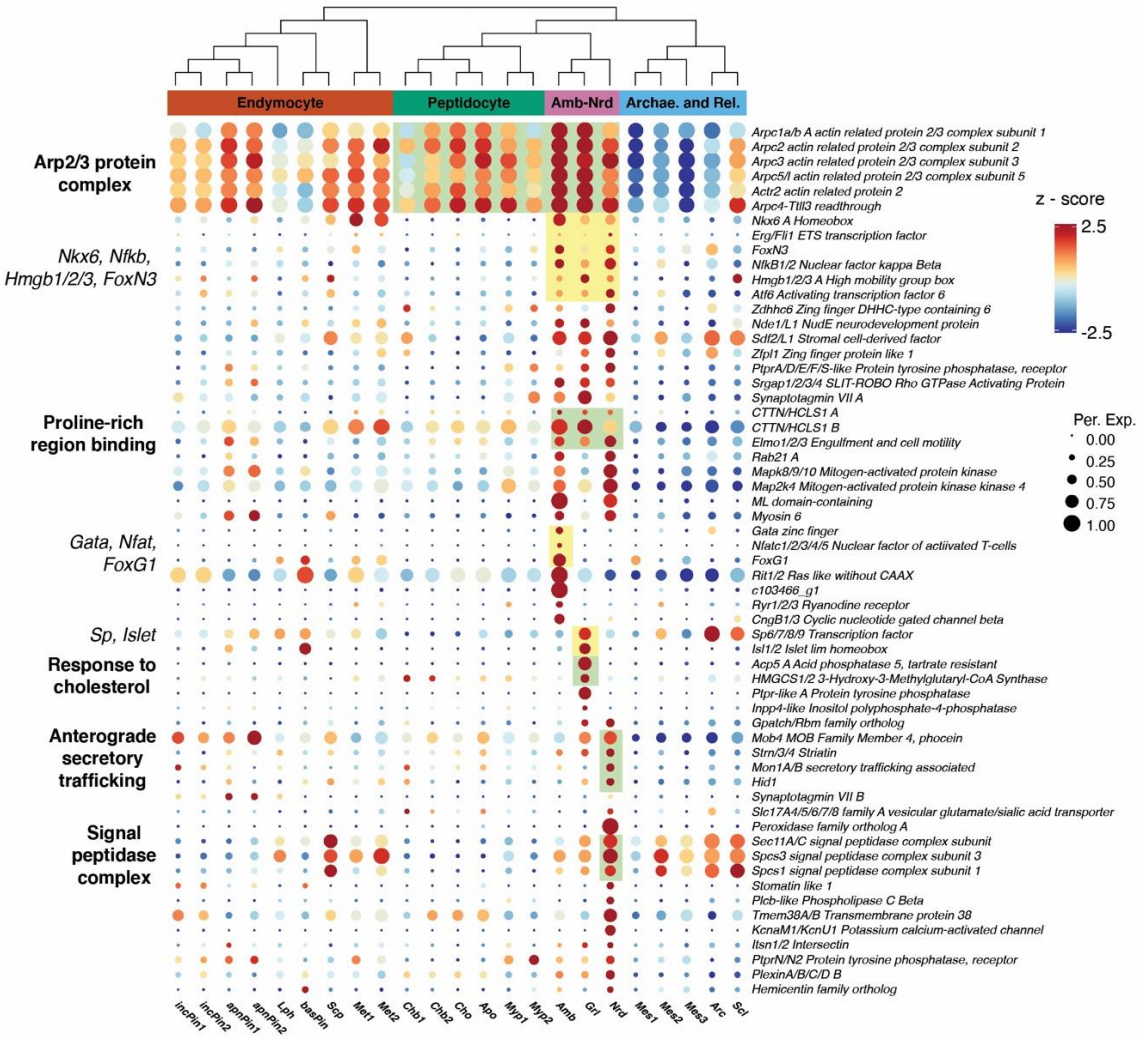

Figure S23. Amoeboid-neuroid family-enriched genes.

Amoeboid-neuroid family-enriched genes. Boxes around dots delineate expression modules of transcription factors (yellow) or functional complexes/pathways (green). Module-specific gene expression in italics, GO terms corresponding to modules in bold.

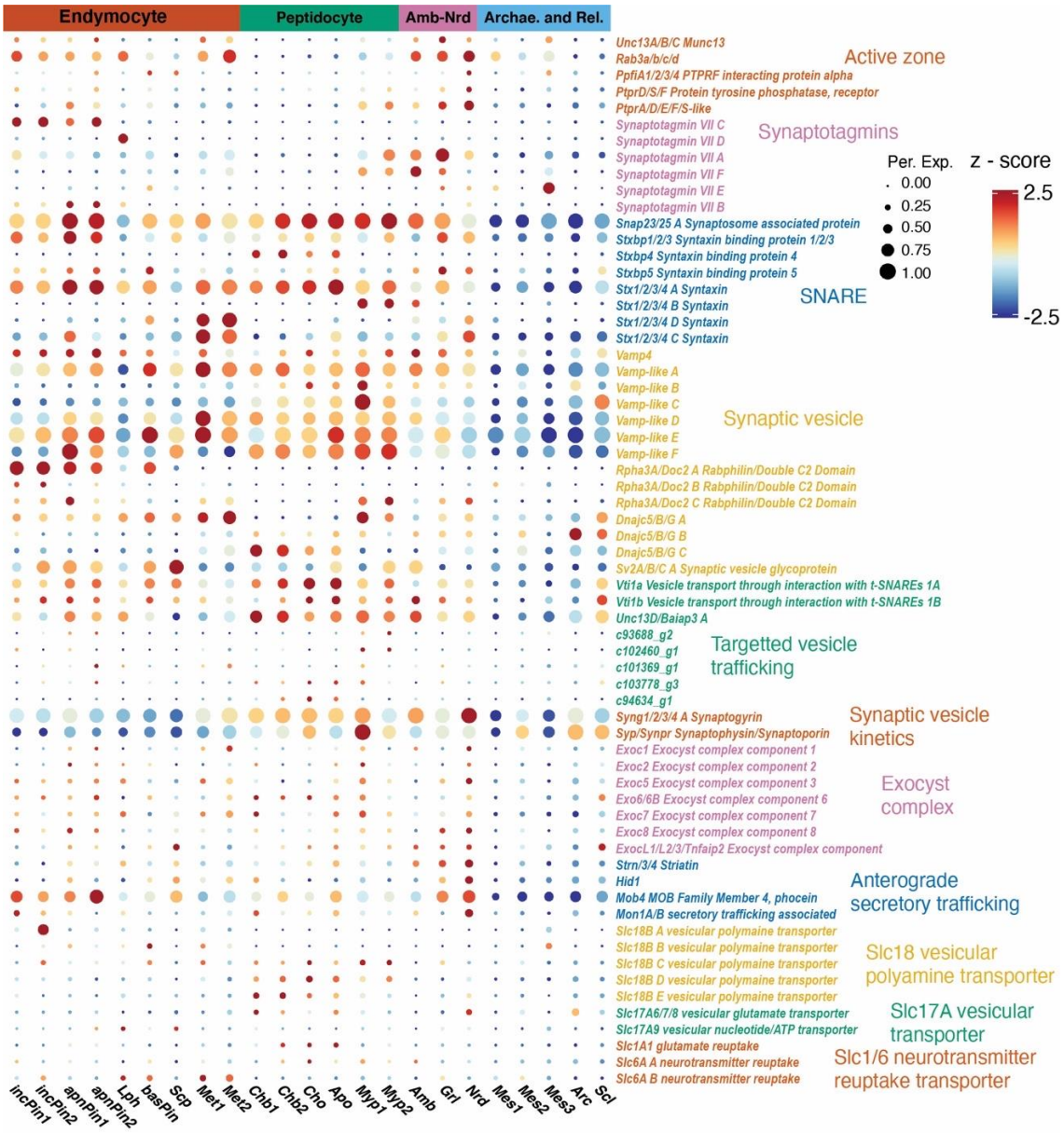

Figure S24. Expression of 'pre'-synaptic/secretory homologs in *Spongilla* cell types.

Dotplots showing scaled expression for orthologs of 'pre'-synaptic/secretory gene homologs in *Spongilla*. Diamater of circle indicates percent of cells within a cell type that are expressing the gene.

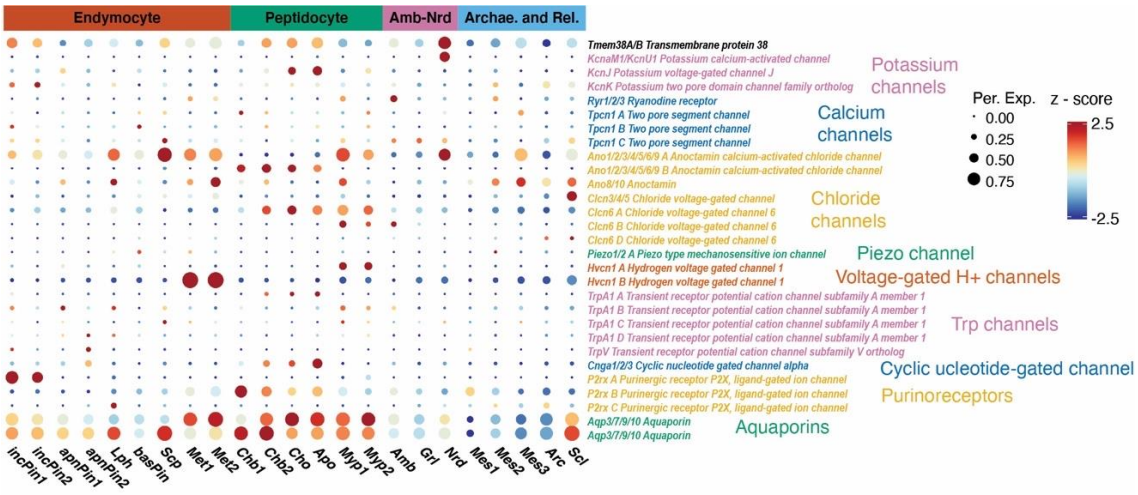

Figure S25. Expression of ion channel homologs in *Spongilla* cell types.

Dotplots showing scaled expression for orthologs of ion channel genes in *Spongilla*. Diameter of circle indicates percent of cells within a cell type that are expressing the gene.

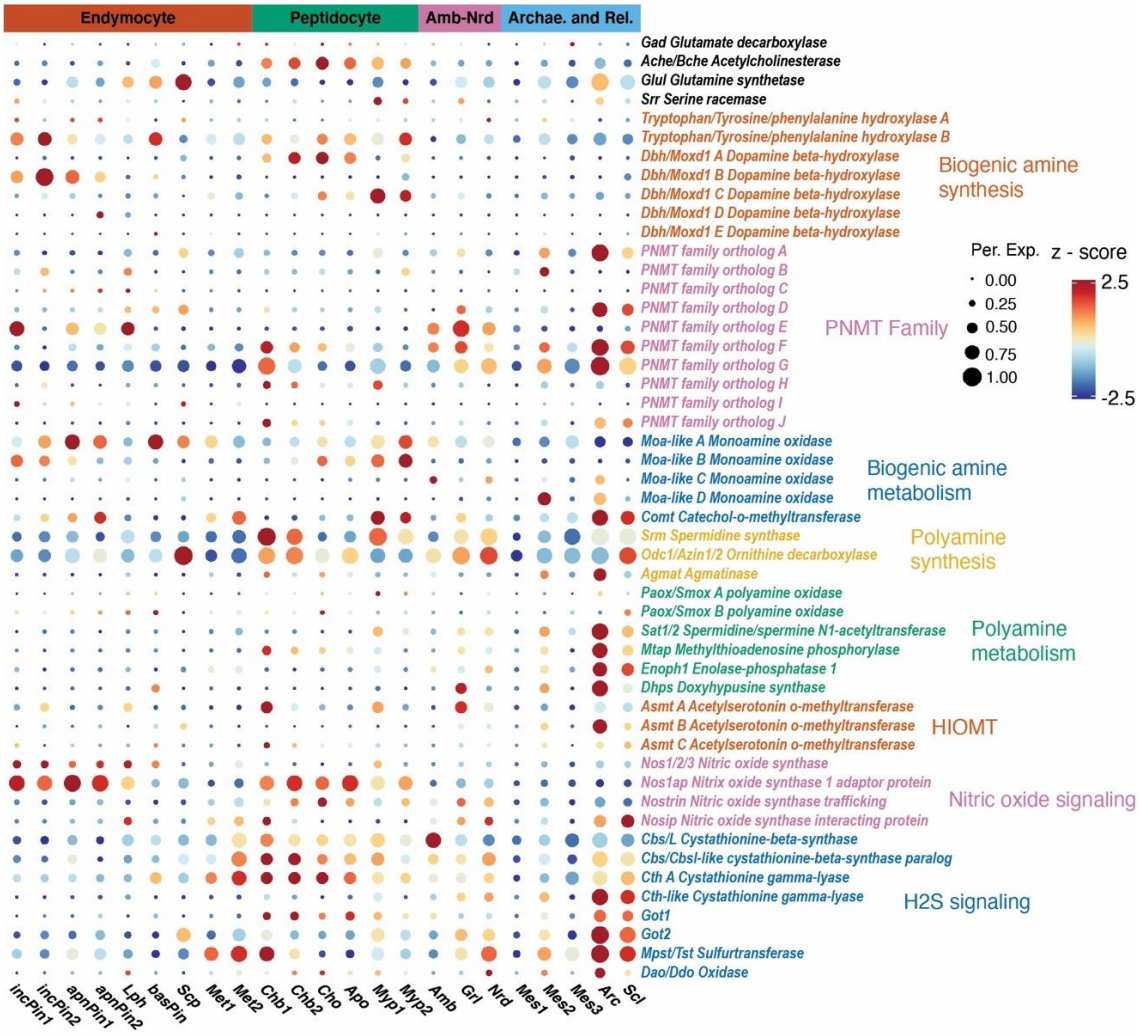

Figure S26. Expression of neurotransmitter pathway homologs in *Spongilla* cell types.
Dotplots showing scaled expression of neurotransmitter pathway synthesis and metabolism gene homologs in *Spongilla*. Diameter of circle indicates percent of cells within a cell type that are expressing the gene.

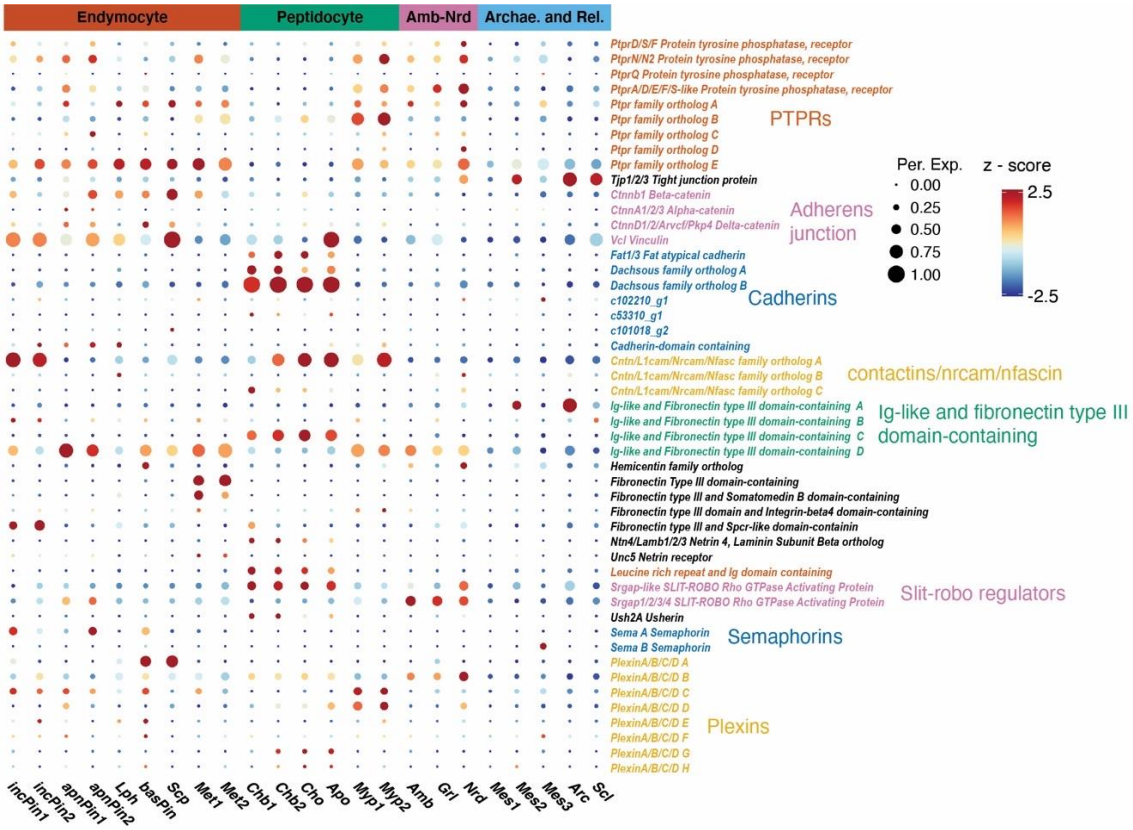

Figure S27. Expression of synapsed-related adhesion homologs in *Spongilla* cell types.

Dotplots showing scaled expression of bilaterian synapse-related adhesion gene homologs in *Spongilla*. Diameter of circle indicates percent of cells within a cell type that are expressing the gene.

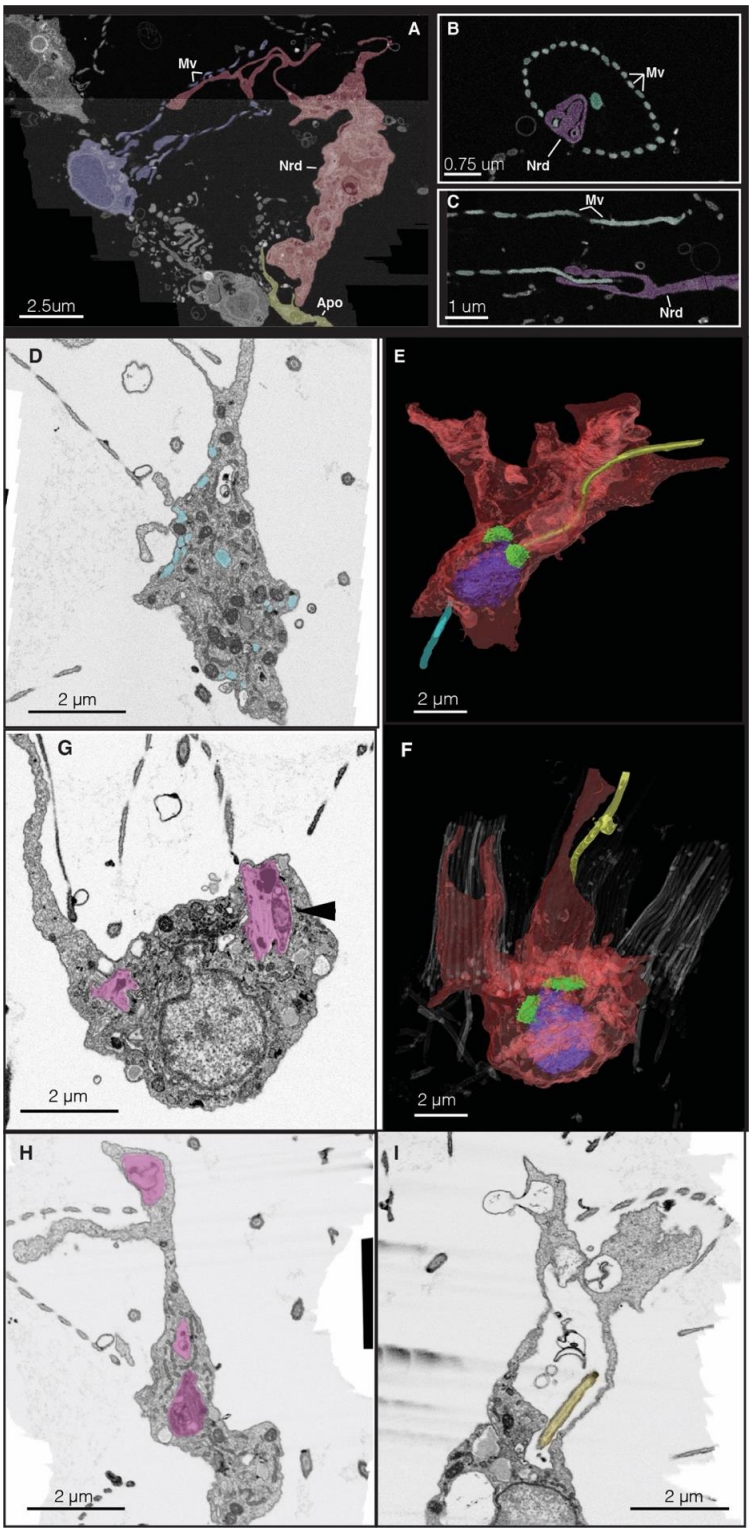

Figure S28. FIB-SEM imaging of neuroid cells in choanocyte chambers.

(A-C) 2D images of neuroid cell filopodia (purple and red) extending into, and enwrapping, choanocyte microvilli (Mv; turquoise and dark blue), and contacting apopylar cell (Apo; yellow). **(D)** FIB-SEM section (5x5x8nm resolution) of neuroid cell two with secretory

vesicles (cyan). **(E)** Rendering of High-resolution neuroid cell two with segmented nuclei (blue), Golgi stacks (green), and choanocyte cilia in close contact with neuroid cell (yellow). **(F)** Rendering of High-resolution neuroid cell one with segmented nuclei (blue), Golgi stacks (green), and choanocyte cilia in close contact with neuroid cell (yellow). Choanocyte microvilli and other cilia rendered in grey. **(G)** High-resolution FIB-SEM section of neuroid cell one with endo/lysosomes or phagosomes (pink overlay) containing likely bacterial cells (arrowhead). **(H)** FIB-SEM section (5x5x8nm resolution) of neuroid cell two with endo/lysosomes or phagosomes (pink). **(I)** FIB-SEM section (5x5x8nm resolution) of neuroid cell two forming a pocket around tip of choanocyte cilia (yellow).

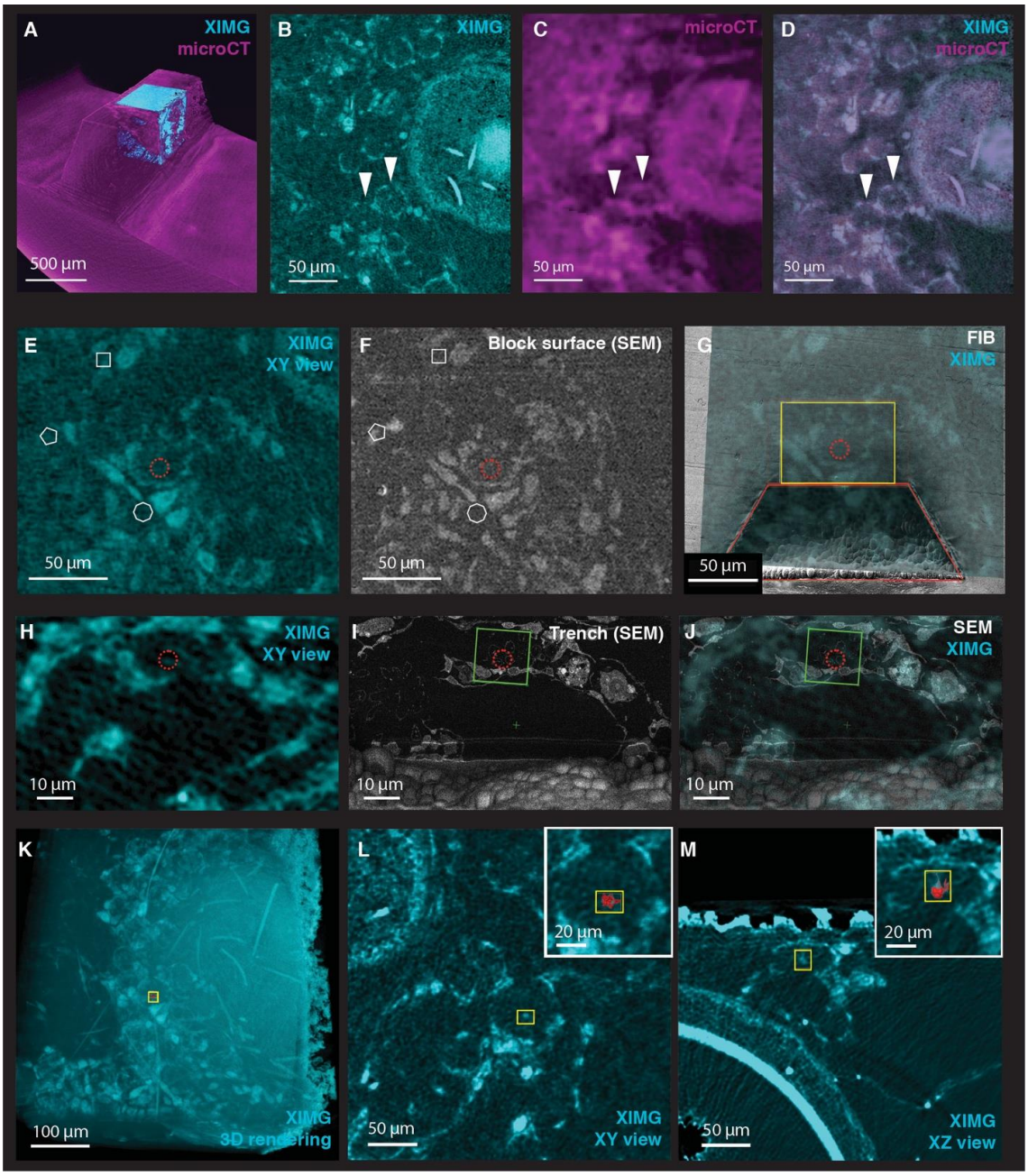

Figure S29. Dual-modality targeting approach for acquiring high-resolution FIB-SEM imaging of neuroid cells in choanocyte chambers.

(A) Rendering of benchtop microCT volume (magenta) registered to a rendering of a smaller high-resolution volume acquired on the EMBL beamline P14 on the PETRA III synchrotron (DESY, Hamburg, Germany) (XIMG, cyan), of the same embedded sponge sample. **(B)** Virtual slice of the XIMG volume parallel to the block surface, approximately 15 µm deep. Arrowheads indicate choanocyte chambers containing a putative neuroid cell, visible at the center of each chamber. These cells were then targeted for FIB-SEM acquisition. **(C)** Virtual slice of the microCT volume at the same plane as shown in B. Arrowheads indicate the same chambers as in B. Note that neuroid cells are not detectable with this instrument. **(D)** Overlay of the virtual

slices shown in B and C. **(E)** XIMG virtual slice showing the predicted cell pattern at the target depth to start FIB-SEM acquisition. The neuroid cells visible in B are located about 11 µm below the plane displayed. **(F)** SEM image of the block surface (ESB detector) after trimming to the predicted depth visualized in E. The white shapes in E and F mark the same positions and are used to register the two images in order to visualize the position exactly at the vertical of the target cell (dotted red circle). **(G)** FIB image (SE detector, greyscale) of the block face after setup of the FIB-SEM run overlaid with the XIMG virtual slice (cyan). Highlighted in yellow the platinum deposition and in red the trench to expose the acquisition surface (orthogonal to this image). The red circle marks the predicted position of the target cell after registration. **(H)** XIMG virtual slice perpendicular to the block surface (i.e. parallel to the orientation of the FIB-SEM stack) at the target position to start the acquisition of the cell of interest. The red circle marks the predicted position of the target cell. **(I)** SEM image (BSE detector) after opening the trench at the predicted position. The red circle marks the predicted position of the cell. The green square is the area that will be acquired during the FIB-SEM run (as seen in Fig. 6, D to E, and fig. S28, D to I). **(J)** Overlay of H and I. **(K)** Overlay of the FIB-SEM volume (bounding box in yellow and segmentation of the target cell in red) with a 3D rendering of the XIMG dataset (cyan). **(L)** Virtual slice of the XIMG dataset parallel to the block surface at the target depth, overlaid with the bounding box of the acquired FIB-SEM volume. The inset shows an enlargement of the same area with the neuroid cell segmentation overlaid. **(M)** Virtual slice of the XIMG dataset parallel to the trench surface at the target depth (XZ view), overlaid with the bounding box of the acquired FIB-SEM volume. The inset shows an enlargement of the same area with the neuroid cell segmentation overlaid.

Cell Type Name	code	Cell Type Family	Name from Weissenfels (1989) (*6*)	Cluster #	Description and explanation of naming
Archaeocytes	Arc	Archaeocytes and relatives	Archäocyten	1	Previously described. Large mesenchymal cells with prominent nucleolus found throughout mesohyl. Principal stem cell in sponges and one of the most abundant cells in the mesohyl.
Sclerocytes	Scl	Archaeocytes and relatives	Skleroblasten	31	Previously described. Elongate mesenchymal cells, often containing growing spicule.
Mesocytes 1	Mes1	Archaeocytes and relatives		33	Newly described cell type in Spongilla lacustris. Medium-sized and irregularly-shaped mesenchymal cell with nucleolus that are relatively uncommon in mesohyl. Name refers to localization in mesohyl and relationship with two other novel mesenchymal cell types.
Mesocytes 2	Mes2	Archaeocytes and relatives		36	Newly described cell type. Medium-sized egg-shaped mesenchymal cells with nucleolus, smaller than archaeocytes. May represent the collencytes referred to by Weissenfels (6). However, the term collencyte has also confusingly been used in other sponges to refer to a type of pinacocyte progenitor, and several authors recommend discontinuing its use (93). Name refers to localization in mesohyl and relationship with two other novel mesenchymal cell types.
Mesocytes 3	Mes3	Archaeocytes and relatives		41	Newly described cell type in Spongilla lacustris. Very rare elongate or fusiform mesohyl cell. Distributional pattern within S. lacustris, beyond presence in the mesohyl, is unclear. Name refers to localization in mesohyl and relationship with two other novel mesenchymal cell types.
Incurrent Pinacocytes1/2	incPin1/2	Endymocytes	Exopinacocyten, Prosendopinacocyten, Porocyten	21-22	Previously described. Epithelial cells forming external tent, outer layer of osculum, and lining subdermal lacunae and incurrent canals. Our single-cell analysis suggests exopinacocytes of the outer tent layer are genetically similar to prosendopinacocytes lining the vestibule and incurrent canals. In contrast, apendopinacocytes lining the excurrent canals and inner layer of the osculum are more distinct. Thus, the traditional term "endopinacocyte" that unites both proesndo- and apendo-pinacocytes appears to be inconsistent with our genetic findings in S. lacustris. Our name reflects the fact that these cells broadly coincide to the incurrent system in sponges.
Apendopinacocytes1/2	apnPin1/2	Endymocytes	Apendopinacocyten	27-28	Previously described. Epithelial cells lining excurrent channels and inner layer of osculum, sometimes ciliated. We observed ciliated cells in both the osculum as well as in excurrent channels near choanocyote apopyles.
Basopinacocytes	basPin	Endymocytes	Exopinacocyten	40	Previously described. Epithelial cells forming layer adherent to substratum, which express spongin short-chain collagen and enwrap mature spicules, cementing them into the base of the sponge. Weissenfels (6) referred to both cells of the basal substratum and cells in the outer layer of the epithelial tent together as "exopinacocytes", which we find as genetically distinct. However, Weissenfels recognized that the cells along the basal substratum were responsible for producing spongin, similar to our findings and those of Nakayama (16). Whether basopinacocytes are synonymous with "spongocytes" described in other sponges to secrete spongin collagen and play a role in gemmule epithelial formation is unclear (93).
Sclerophorocytes	Scp	Endymocytes	Begleitzellen	37	Previously described. Mesenchymal cells typically observed in contact with mature spicules, which function in spicule transport and skeletogenesis and described as “transport cells” in (16). Here, we propose a new Greek translation that honors previous names and recognizes their role in transporting spicules.
Lophocytes	Lph	Endymocytes	Lophocyten	18	Previously described. Migratory cells with filopodia and elongate, asymmetric shape. Most prevalent near basopinacocytes and incurrent pinacocytes. Described across porifera as being important in collagen synthesis (93), consistent with our findings that they upregulate distinct collagens.
Metabaloyctes1/2	Met1/2	Endymocytes		24-25	Newly described cell type. Large, multipolar mesohyl cells lacking nucleolus. Relatively uncommon in mesohyl, although often found in proximity to baso- and incurrent pinacocytes. This cell type does not clearly correspond to any previously-defined morphological cell type in freshwater sponges, and its homology to mesenchymal cell types in other sponges is unclear. The name was chosen to reflect their apparent functional role as metabolic cells.
Choanocblasts1/2	Chb1/2	Peptidocytes	Choanoblasten	8-9	Previously described. Proliferative medium-size cells lacking collar and cilium that are progenitors of choanocytes and apopylar cels. Often found in the mesohyl in close proximity to choanocyte chambers.

Choanocytes	Cho	Peptidocytes	Choanocyten	10	Previously described. Mature choanocytes, small cells with collar and cilium that play a central role in feeding and driving water flow.
Apopylar Cells	Apo	Peptidocytes	Konuszellen	11	Previously described. Cells forming the excurrent pore of mature choanocyte chamber. Cilium present but lacks microvilli collar.
Myopeptidocytes1/2	Myp1/2	Peptidocytes		4-5	Newly described cell type. Medium-sized mesechymal cells with prominent vacuoles and long projections forming cellular network, nucleolus absent. Their name reflects their expression profile, which combines both muscle contractile genes and genes involved in phagocytosis and digestion.
Amoebocytes	Amb	Amoeboid-Neuroid	Amöbocyten	39	Previously described. Small mesenchymal cells lacking nucleolus, observed engulfing other cells, similar to that noted by Weissenfels (6).
Neuroid	Nrd	Amoeboid-Neuroid		35	Previously described. Small, multipolar mesenchymal cells with thin projections, usually found in choanocyte chambers. Similar to the neuroid cells described by Pavans de Ceccatty (23), which he later termed “neuroid” cells (94). Also synonymous with central cells described by Reiswig and Brown (24). We use "neuroid" cell as it predates the term central cell and is a better functional description of these cells.
Granulocytes	Grl	Amoeboid-Neuroid	Granulocyten	30	Previously described. Small mesenchymal cells with projections, scattered in mesenchyme but also often found in the tent and in proximity to pinacocytes.
			Trophocyten		Nutrient cells present during gemmule formation, not sampled in our study.
			Thesocyten (statocyten)		Cells constituting the gemmule, not sampled in our study.
			Spermatogonien		Sperm cells present during sexual maturity, not sampled in our study.
			Oogonien		Egg cells present during sexual maturity, not sampled in our study.

Table S1. Cell types in *Spongilla lacustris*, their morphology and distribution.

References and Notes

1. M. Srivastava, O. Simakov, J. Chapman, B. Fahey, M. E. A. Gauthier, T. Mitros, G. S. Richards, C. Conaco, M. Dacre, U. Hellsten, C. Larroux, N. H. Putnam, M. Stanke, M. Adamska, A. Darling, S. M. Degnan, T. H. Oakley, D. C. Plachetzki, Y. Zhai, M. Adamski, A. Calcino, S. F. Cummins, D. M. Goodstein, C. Harris, D. J. Jackson, S. P. Leys, S. Shu, B. J. Woodcroft, M. Vervoort, K. S. Kosik, G. Manning, B. M. Degnan, D. S. Rokhsar, The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010). doi:10.1038/nature09201 Medline

2. G. R. D. Elliott, S. P. Leys, Coordinated contractions effectively expel water from the aquiferous system of a freshwater sponge. *J. Exp. Biol.* **210**, 3736–3748 (2007). doi:10.1242/jeb.003392 Medline

3. A. Alié, T. Hayashi, I. Sugimura, M. Manuel, W. Sugano, A. Mano, N. Satoh, K. Agata, N. Funayama, The ancestral gene repertoire of animal stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7093–E7100 (2015). doi:10.1073/pnas.1514789112 Medline

4. F. A. Wolf, F. K. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L. Simon, F. J. Theis, PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59–9 (2019). doi:10.1186/s13059-019-1663-x Medline

5. J. F. Peña, A. Alié, D. J. Richter, L. Wang, N. Funayama, S. A. Nichols, Conserved expression of vertebrate microvillar gene homologs in choanocytes of freshwater sponges. *Evodevo* **7**, 13 (2016). doi:10.1186/s13227-016-0050-x Medline

6. N. Weissenfels, *Biologie und Mikroskopische Anatomie der Süßwasserschwämme (Spongillidae)* (Gustav Fischer, 1989).

7. C. Liang, FANTOM Consortium, A. R. Forrest, G. P. Wagner, The statistical geometry of transcriptome divergence in cell-type evolution and cancer. *Nat. Commun.* **6**, 6066 (2015). doi:10.1038/ncomms7066 Medline

8. A. J. Tarashansky, J. M. Musser, M. Khariton, P. Li, D. Arendt, S. R. Quake, B. Wang, Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021). doi:10.7554/eLife.66747 Medline

9. A. Sebé-Pedrós, E. Chomsky, K. Pang, D. Lara-Astiaso, F. Gaiti, Z. Mukamel, I. Amit, A. Hejnol, B. M. Degnan, A. Tanay, Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 (2018). doi:10.1038/s41559-018-0575-6 Medline

10. M. Nickel, C. Scheer, J. U. Hammel, J. Herzen, F. Beckmann, The contractile sponge epithelium sensu lato—Body contraction of the demosponge Tethya wilhelma is mediated by the pinacoderm. *J. Exp. Biol.* **214**, 1692–1698 (2011). doi:10.1242/jeb.049148 Medline

11. J. M. Miano, X. Long, K. Fujiwara, Serum response factor: Master regulator of the actin cytoskeleton and contractile apparatus. *Am. J. Physiol. Cell Physiol.* **292**, C70–C81 (2007). doi:10.1152/ajpcell.00386.2006 Medline

12. K. Ellwanger, A. Eich, M. Nickel, GABA and glutamate specifically induce contractions in the sponge Tethya wilhelma. *J. Comp. Physiol.* **193**, 1–11 (2007). doi:10.1007/s00359-006-0165-y Medline

13. G. R. D. Elliott, S. P. Leys, Evidence for glutamate, GABA and NO in coordinating behaviour in the sponge, Ephydatia muelleri (Demospongiae, Spongillidae). *J. Exp. Biol.* **213**, 2310–2321 (2010). doi:10.1242/jeb.039859 Medline

14. K. Ellwanger, M. Nickel, Neuroactive substances specifically modulate rhythmic body contractions in the nerveless metazoon Tethya wilhelma (Demospongiae, Porifera). *Front. Zool.* **3**, 7 (2006). doi:10.1186/1742-9994-3-7 Medline

15. A. Rivera, I. Winters, A. Rued, S. Ding, D. Posfai, B. Cieniewicz, K. Cameron, L. Gentile, A. Hill, The evolution and function of the Pax/Six regulatory network in sponges. *Evol. Dev.* **15**, 186–196 (2013). doi:10.1111/ede.12032 Medline

16. S. Nakayama, K. Arima, K. Kawai, K. Mohri, C. Inui, W. Sugano, H. Koba, K. Tamada, Y. J. Nakata, K. Kishimoto, M. Arai-Shindo, C. Kojima, T. Matsumoto, T. Fujimori, K. Agata, N. Funayama, Dynamic Transport and Cementation of Skeletal Elements Build Up the Pole-and-Beam Structured Skeleton of Sponges. *Curr. Biol.* **25**, 2549–2554 (2015). doi:10.1016/j.cub.2015.08.023 Medline

17. T. Kawahara, M. T. Quinn, J. D. Lambeth, Molecular evolution of the reactive oxygen-generating NADPH oxidase (Nox/Duox) family of enzymes. *BMC Evol. Biol.* **7**, 109–121 (2007). doi:10.1186/1471-2148-7-109 Medline

18. M. Carrasco-Rando, M. Ruiz-Gómez, Mind bomb 2, a founder myoblast-specific protein, regulates myoblast fusion and muscle stability. *Development* **135**, 849–857 (2008). doi:10.1242/dev.015529 Medline

19. D. Laundon, B. T. Larson, K. McDonald, N. King, P. Burkhardt, The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLOS Biol.* **17**, e3000226–e22 (2019). doi:10.1371/journal.pbio.3000226 Medline

20. V. Gopalan, A. Singh, F. Rashidi Mehrabadi, L. Wang, E. Ruppin, H. E. Arda, S. Hannenhalli, A Transcriptionally Distinct Subpopulation of Healthy Acinar Cells Exhibit Features of Pancreatic Progenitors and PDAC. *Cancer Res.* **81**, 3958–3970 (2021). doi:10.1158/0008-5472.CAN-21-0427 Medline

21. J. T. H. Wang, M. C. Kerr, S. Karunaratne, A. Jeanes, A. S. Yap, R. D. Teasdale, The SNX-PX-BAR family in macropinocytosis: The regulation of macropinosome formation by SNX-PX-BAR proteins. *PLOS ONE* **5**, e13763–e13 (2010). doi:10.1371/journal.pone.0013763 Medline

22. D. Arendt, The Evolutionary Assembly of Neuronal Machinery, The Evolutionary Assembly of Neuronal Machinery. *Curr. Biol.* **30**, R603–R616 (2020). doi:10.1016/j.cub.2020.04.008 Medline

23. M. P. de Ceccatty, Le système nerveux des éponges calcaires et siliceuses. *Ann. Sci. Nat. Zool. Biol. Anim.* **17**, 203–288 (1955).

24. H. M. Reiswig, M. J. Brown, The central cells of sponges. *Zoomorphologie* **88**, 81–94 (1977). doi:10.1007/BF01880648

25. V. Hartenstein, L. Mandal, The blood/vascular system in a phylogenetic perspective. *BioEssays* **28**, 1203–1210 (2006). doi:10.1002/bies.20497 Medline

26. T. L. Gumienny, E. Brugnera, A.-C. Tosello-Trampont, J. M. Kinchen, L. B. Haney, K. Nishiwaki, S. F. Walk, M. E. Nemergut, I. G. Macara, R. Francis, T. Schedl, Y. Qin, L. Van Aelst, M. O. Hengartner, K. S. Ravichandran, CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* **107**, 27–41 (2001). doi:10.1016/S0092-8674(01)00520-7 Medline

27. T. Khurana, J. A. Brzostowski, A. R. Kimmel, A Rab21/LIM-only/CH-LIM complex regulates phagocytosis via both activating and inhibitory mechanisms. *EMBO J.* **24**, 2254–2264 (2005). doi:10.1038/sj.emboj.7600716 Medline

28. D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019). doi:10.1016/j.conb.2019.01.022 Medline

29. R. M. J. Palmer, A. G. Ferrige, S. Moncada, Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor. *Nature* **327**, 524–526 (1987). doi:10.1038/327524a0 Medline

30. S. A. V. Fortunato, M. Adamski, O. M. Ramos, S. Leininger, J. Liu, D. E. K. Ferrier, M. Adamska, Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. *Nature* **514**, 620–623 (2014). doi:10.1038/nature13881 Medline

31. Y. Li, D. Zhao, T. Horie, G. Chen, H. Bao, S. Chen, W. Liu, R. Horie, T. Liang, B. Dong, Q. Feng, Q. Tao, X. Liu, Conserved gene regulatory module specifies lateral neural borders across bilaterians. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6352–E6360 (2017). doi:10.1073/pnas.1704194114 Medline

32. S. M. Bell, L. Zhang, Y. Xu, V. Besnard, S. E. Wert, N. Shroyer, J. A. Whitsett, Kruppel-like factor 5 controls villus formation and initiation of cytodifferentiation in the embryonic intestinal epithelium. *Dev. Biol.* **375**, 128–139 (2013). doi:10.1016/j.ydbio.2012.12.010 Medline

33. P. R. H. Steinmetz, A. Aman, J. E. M. Kraus, U. Technau, Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. *Nat. Ecol. Evol.* **1**, 1535–1542 (2017). doi:10.1038/s41559-017-0285-5 Medline

34. D. Arendt, E. Benito-Gutierrez, T. Brunet, H. Marlow, Gastric pouches and the mucociliary sole: Setting the stage for nervous system evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20150286–20150286 (2015). doi:10.1098/rstb.2015.0286 Medline

35. L. L. Moroz, A. B. Kohn, Independent origins of neurons and synapses: Insights from ctenophores. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150041 (2016). doi:10.1098/rstb.2015.0041 Medline

36. D. Arendt, Elementary nervous systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200347 (2021). doi:10.1098/rstb.2020.0347 Medline

37. J. Musser, Gitlab repository zip for Profiling cellular diversity in sponges informs animal cell type and nervous system evolution, Zenodo (2021); doi:10.5281/zenodo.5094890

38. N. Funayama, M. Nakatsukasa, S. Kuraku, K. Takechi, M. Dohi, N. Iwabe, T. Miyata, K. Agata, Isolation of Ef silicatein and Ef lectin as molecular markers for sclerocytes and

cells involved in innate immunity in the freshwater sponge Ephydatia fluviatilis. *Zool. Sci.* **22**, 1113–1122 (2005). doi:10.2108/zsj.22.1113 Medline

39. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). doi:10.1038/nbt.1883 Medline

40. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). doi:10.1093/bioinformatics/btv351 Medline

41. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016). doi:10.1093/molbev/msw046 Medline

42. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2011). doi:10.1038/msb.2011.75 Medline

43. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi:10.1093/bioinformatics/btp348 Medline

44. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033 Medline

45. P. Lapébie, E. Gazave, A. Ereskovsky, R. Derelle, C. Bézac, E. Renard, E. Houliston, C. Borchiellini, WNT/β-catenin signalling and epithelial patterning in the homoscleromorph sponge Oscarella. *PLOS ONE* **4**, e5823 (2009). doi:10.1371/journal.pone.0005823 Medline

46. S. A. Fortunato, S. Leininger, M. Adamska, Evolution of the Pax-Six-Eya-Dach network: The calcisponge case study. *Evodevo* **5**, 23–12 (2014). doi:10.1186/2041-9139-5-23 Medline

47. A. Riesgo, N. Farrar, P. J. Windsor, G. Giribet, S. P. Leys, The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol. Biol. Evol.* **31**, 1102–1120 (2014). doi:10.1093/molbev/msu057 Medline

48. S. A. Nichols, W. Dirks, J. S. Pearse, N. King, Early evolution of animal cell signaling and adhesion genes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12451–12456 (2006). doi:10.1073/pnas.0604065103 Medline

49. J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork, Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017). doi:10.1093/molbev/msx148 Medline

50. M. Blum, H. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G. A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, A. Bateman, R. D. Finn, The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021). doi:10.1093/nar/gkaa977 Medline

51. N.-P. D. Nguyen, S. Mirarab, K. Kumar, T. Warnow, Ultra-large alignments using phylogeny-aware profiles. *Genome Biol.* **16**, 124 (2015). doi:10.1186/s13059-015-0688-z Medline

52. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi:10.1038/nbt.4096 Medline

53. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015). doi:10.1016/j.cell.2015.05.002 Medline

54. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi:10.1186/1471-2105-9-559 Medline

55. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15–5 (2018). doi:10.1186/s13059-017-1382-0 Medline

56. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015). doi:10.1093/bioinformatics/btv325 Medline

57. R. G. Parra, N. Papadopoulos, L. Ahumada-Arranz, J. E. Kholtei, N. Mottelson, Y. Horokhovsky, B. Treutlein, J. Soeding, Reconstructing complex lineage trees from scRNA-seq data using MERLoT. *Nucleic Acids Res.* **47**, 8961–8974 (2019). doi:10.1093/nar/gkz706 Medline

58. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006). doi:10.1093/molbev/msj030 Medline

59. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W425 (2016). doi:10.1093/nar/gkw290 Medline

60. Y. Benjamini, R. Heller, D. Yekutieli, Selective inference in complex research, Selective inference in complex research. *Philos. Trans. A Math. Phys. Eng. Sci.* **367**, 4255–4271 (2009). Medline

61. J. Clavel, G. Escarguel, G. Merceron, mvmorph: An r package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015). doi:10.1111/2041-210X.12420

62. T. C. Südhof, The presynaptic active zone. *Neuron* **75**, 11–25 (2012). doi:10.1016/j.neuron.2012.06.012 Medline

63. M. Missler, T. C. Südhof, T. Biederer, Synaptic cell adhesion. *Cold Spring Harb. Perspect. Biol.* **4**, a005694 (2012). Medline

64. L. L. Moroz, A. B. Kohn, Unbiased View of Synaptic and Neuronal Gene Complement in Ctenophores: Are There Pan-neuronal and Pan-synaptic Genes across Metazoa? *Integr. Comp. Biol.* **55**, 1028–1049 (2015). doi:10.1093/icb/icv104 Medline

65. B. Olas, Hydrogen sulfide in signaling pathways. *Clin. Chim. Acta* **439**, 212–218 (2015). doi:10.1016/j.cca.2014.10.037 Medline

66. H. O. Lawal, D. E. Krantz, SLC18: Vesicular neurotransmitter transporters for monoamines and acetylcholine. *Mol. Aspects Med.* **34**, 360–372 (2013). doi:10.1016/j.mam.2012.07.005 Medline

67. B. Binotti, R. Jahn, J. J. E. Chua, Functions of Rab Proteins at Presynaptic Sites. *Cells* **5**, 7 (2016). doi:10.3390/cells5010007 Medline

68. A. A. Alabi, R. W. Tsien, Synaptic vesicle pools and dynamics. *Cold Spring Harb. Perspect. Biol.* **4**, a013680 (2012). doi:10.1101/cshperspect.a013680 Medline

69. T. C. Südhof, J. Rizo, Synaptic vesicle exocytosis. *Cold Spring Harb. Perspect. Biol.* **3**, a005637–a005637 (2011). doi:10.1101/cshperspect.a005637 Medline

70. Y. Saheki, P. De Camilli, Synaptic vesicle endocytosis. *Cold Spring Harb. Perspect. Biol.* **4**, a005645 (2012). Medline

71. M. Sheng, E. Kim, The Postsynaptic Organization of Synapses, *Cold Spring Harb. Perspect. Biol.* **3**, a005678 (2011).

72. I. Gaspar, F. Wippich, A. Ephrussi, Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017). doi:10.1261/rna.061184.117 Medline

73. K. J. Schippers, S. A. Nichols, Evidence of Signaling and Adhesion Roles for β-Catenin in the Sponge Ephydatia muelleri. *Mol. Biol. Evol.* **35**, 1407–1421 (2018). doi:10.1093/molbev/msy033 Medline

74. M. Nickel, Kinetics and rhythm of body contractions in the sponge Tethya wilhelma (Porifera: Demospongiae). *J. Exp. Biol.* **207**, 4515–4524 (2004). doi:10.1242/jeb.01289 Medline

75. M. Nickel, F. Brümmer, Body extension types of Tethya wilhelma: Cellular organisation and their locomotory function. *Boll. Musei Ist. Biol. Univ. Genova* **68**, 483–489 (2004).

76. M. C. Willingham, A. V. Rutherford, The use of osmium-thiocarbohydrazide-osmium (OTO) and ferrocyanide-reduced osmium methods to enhance membrane contrast and preservation in cultured cells. *J. Histochem. Cytochem.* **32**, 455–460 (1984). doi:10.1177/32.4.6323574 Medline

77. J. C. Tapia, N. Kasthuri, K. J. Hayworth, R. Schalek, J. W. Lichtman, S. J. Smith, J. Buchanan, High-contrast en bloc staining of neuronal tissue for field emission scanning electron microscopy. *Nat. Protoc.* **7**, 193–206 (2012). doi:10.1038/nprot.2011.439 Medline

78. N. L. Schieber, S. J. Nixon, R. I. Webb, V. M. J. Oorschot, R. G. Parton, Modern Approaches for Ultrastructural Analysis of the Zebrafish Embryo. *Methods Cell Biol.* **96**, 425–442 (2010). doi:10.1016/S0091-679X(10)96018-4

79. F. Cipriani, F. Felisaz, L. Launer, J.-S. Aksoy, H. Caserotto, S. Cusack, M. Dallery, F. di-Chiaro, M. Guijarro, J. Huet, S. Larsen, M. Lentini, J. McCarthy, S. McSweeney, R. Ravelli, M. Renier, C. Taffut, A. Thompson, G. A. Leonard, M. A. Walsh, Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1251–1259 (2006). doi:10.1107/S0907444906030587 Medline

80. M. Polikarpov, G. Bourenkov, I. Snigireva, A. Snigirev, S. Zimmermann, K. Csanko, S. Brockhauser, T. R. Schneider, Visualization of protein crystals by high-energy phase-contrast X-ray imaging. *Acta Crystallogr. D Struct. Biol.* **75**, 947–958 (2019). doi:10.1107/S2059798319011379 Medline

81. S. Zabler, P. Cloetens, J. P. Guigay, J. Baruchel, M. Schlenker, Optimization of phase contrast imaging using hard x rays. *Rev. Sci. Instrum.* **76**, 073705 (2005). doi:10.1063/1.1960797

82. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004). doi:10.1109/TIP.2003.819861 Medline

83. P. Cloetens, W. Ludwig, J. Baruchel, D. Van Dyck, J. Van Landuyt, J. P. Guigay, M. Schlenker, Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *Appl. Phys. Lett.* **75**, 2912–2914 (1999). doi:10.1063/1.125225

84. D. Gürsoy, F. De Carlo, X. Xiao, C. Jacobsen, TomoPy: A framework for the analysis of synchrotron tomographic data. *J. Synchrotron Radiat.* **21**, 1188–1193 (2014). doi:10.1107/S1600577514013939 Medline

85. A. Limaye, Drishti: A volume exploration and presentation tool, in *Proc. SPIE* 8506 (2012), p. 85060X.

86. A. Horňáková, J. H. Lange, B. Andres, Analysis and optimization of graph decompositions by lifted multicuts, in *Proceedings of the the 34th International Conference on Machine Learning* (2017), vol. 70, pp. 1539–1548.

87. C. Pape, A. Matskevych, A. Wolny, J. Hennies, G. Mizzon, M. Louveaux, J. Musser, A. Maizel, D. Arendt, A. Kreshuk, Leveraging Domain Knowledge to Improve Microscopy Image Segmentation With Lifted Multicuts. *Front. Comput. Sci.* **1**, 6 (2019). doi:10.3389/fcomp.2019.00006

88. C. Sommer, C. Straehle, U. Köthe, F. A. Hamprecht, Ilastik: Interactive learning and segmentation toolkit, in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, 2011), pp. 230–233.

89. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001). doi:10.1023/A:1010933404324

90. T. Beier, C. Pape, N. Rahaman, T. Prange, S. Berg, D. D. Bock, A. Cardona, G. W. Knott, S. M. Plaza, L. K. Scheffer, U. Koethe, A. Kreshuk, F. A. Hamprecht, Multicut brings

automated neurite segmentation closer to human performance. *Nat. Methods* **14**, 101–102 (2017). doi:10.1038/nmeth.4151 Medline

91. C. Pape, T. Beier, P. Li, V. Jain, D. D. Bock, A. Kreshuk, Solving large multicut problems for connectomics via domain decomposition, in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2017), pp. 1–10.

92. H. M. Vergara, C. Pape, K. I. Meechan, V. Zinchenko, C. Genoud, A. A. Wanner, K. N. Mutemi, B. Titze, R. M. Templin, P. Y. Bertucci, O. Simakov, W. Dürichen, P. Machado, E. L. Savage, L. Schermelleh, Y. Schwab, R. W. Friedrich, A. Kreshuk, C. Tischer, D. Arendt, Whole-body integration of gene expression and single-cell morphology. *Cell* **184**, 4819–4837.e22 (2021). doi:10.1016/j.cell.2021.07.017 Medline

93. T. L. Simpson, *The Cell Biology of Sponges* (Springer, 1984).

94. M. P. de Ceccatty, The origin of the integrative systems: A change in view derived from research on coelenterates and sponges. *Perspect. Biol. Med.* **17**, 379–391 (1974). doi:10.1353/pbm.1974.0060 Medline