Changes of cell-type diversity in the polyp-to-medusa metagenesis of the scyphozoan jellyfish *Aurelia coerulea* (formerly sp.1)

Oliver Link[1], Stefan M. Jahnel[2], Kristin Janicek[1], Johanna Kraus[1], Juan Daniel Montenegro[1], Bob Zimmerman[1], Brittney Wick[3], Alison G. Cole[1]*, and Ulrich Technau[1,4]*

[1] Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, Djerassiplatz 1, 1030 Vienna.

[2] Institute of Molecular Biotechnology, Dr.-Bohr-Gasse 3, 1030 Vienna.

[3] UCSC Cellbrowser, University of California, Santa Cruz, USA.

[4] Research platform Single Cell Regulation of Stem Cells, University of Vienna, Djerassiplatz 1, 1030 Vienna.

*Corresponding authors: alison.cole@univie.ac.at and ulrich.technau@univie.ac.at

Keywords:

Aurelia, Scyphozoa, strobilation, single cell RNAseq, developmental cell atlas, life cycle, jellyfish

Abstract

The life cycle of most medusozoan cnidarians is marked by the metagenesis from the asexually reproducing sessile polyp and the sexually reproducing motile medusa. At present it is unknown to what extent this drastic morphological transformation is accompanied by changes in the cell type composition. Here, we provide a single cell transcriptome atlas of the cosmopolitan scyphozoan *Aurelia coerulea* focussing on changes in cell-type composition during the transition from polyp to medusa. Notably, this transition marked by an increase in cell type diversity, including an expansion of neural subtypes. We find that two families of neuronal lineages are specified by homologous transcription factors in the sea anemone *Nematostella vectensis* and *Aurelia coerulea*, suggesting an origin in the common ancestor of medusozoans and anthozoans about 500 Myr ago. Our analysis suggests that gene duplications might be drivers for the increase of cellular complexity during the evolution of cnidarian neuroglandular lineages. One key medusozoan-specific cell type is the striated muscle in the subumbrella. Analysis of muscle fiber anatomy and gene expression raises the possibility that the striated muscles arise from a population of smooth muscle cells during strobilation. Although smooth and striated muscles are phenotypically distinct, both have a similar contractile complex, in contrast to bilaterian smooth and striated muscles. This suggests that in *Aurelia*, smooth and striated muscle cells may derive from the same progenitor cells.

Teaser

Single cell transcriptome atlas across the jellyfish life cycle reveals emergence of novel medusa-specific cell types is associated with expression of medusa-specific paralogs.

Introduction

Cell-type diversification is a hallmark in the evolution of metazoans, and the principles that underlie these processes are still barely understood. Advances in single cell sequencing allow the cataloguing of differential transcriptome usage across different cell types from entire organisms (reviewed in (1,2)). Comparative methods for such data across phylogenetic distances remain a non-trivial task, and robust phylogenetic sampling will be required to make any clear evolutionary inferences from cross-organism comparisons. To understand the evolution of cell-type diversification that coincided with the emergence of complex body plans in the animal kingdom, members of basal branching animal phylum Cnidaria hold the key to reconstructing ancestral cell-type states present during the explosion of animal morphotypes that remain in modern day. Cnidarians are the phylogenetic sister group to Bilateria (3–5),

which comprise the vast majority of all animal phyla. In this regard understanding the phylum Cnidaria is key for the identification of key bilaterian traits.

The phylum Cnidaria is composed of the subphylum Anthozoa, which encompasses sea anemones and corals, and its sister group, the Medusozoa. Medusozoa differ from the anthozoans in that these lineages, except for Staurozoa (6), possess a triphasic life cycle that includes a free-swimming larva, sessile polyp, and a free-swimming jellyfish (a.k.a. medusa) stage in their life-histories (6,7). The drastic metagenesis from the polyp into free-swimming jellyfish in different medusozoan clades provides opportunities to investigate the evolutionary principles underlying the development of any new cell types that accompany the switch from a sedentary to a free-living lifestyle. One such cell type is the cnidarian striated muscle, which morphologically has only been identified in the medusozoan jellyfish (8). Additionally, a complex sensory organ, the rhopalia, and a specific set of motor neurons comprising the motor nerve net (MNN), develop during the polyp-medusa metamorphosis (8–10). In scyphozoans like *Aurelia,* metagenesis occurs via polydisc strobilation, an asexual horizontal fission process along the oral-aboral axis of the polyp which enables the organism to produce many free-swimming individuals at once (11).

In this study, we provide a molecular foundation for investigating mechanisms of cell type diversification in Scyphozoa by generating a cell-type atlas of a cosmopolitan scyphozoan model, *Aurelia coerulea* (formerly sp.1). We used single-cell RNA sequencing to produce whole animal-, as well as tissue specific-, single-cell RNAseq libraries that cover the transition from polyp (also called scyphistoma) to jellyfish (a.k.a. medusa). We characterise the composition of identified cell type families across the life cycle and identify cell types that define the jellyfish as free-swimming individuals, including various neuronal populations and striated muscles, confirming anatomical reports (8,12). We found parallels with anthozoans in terms of reported cellular complexity with regards to neural composition and cnidocyte specification pathways (13–15). The medusa-specific striated muscles enable jellyfish to propel through the water column and allow for a highly efficient swimming ability (16). We characterise the molecular signatures of the striated muscles compared to the smooth muscles of the polyp and identified a set of genes that is shared between both muscle phenotypes. These genes correspond to a shared contractile complex between both muscle types. With this analysis we set the foundation for comparative studies addressing the assembly of the striated muscle phenotype in Cnidaria.

Results

The typical life cycle of medusozoans involves the transition from a sessile polyp (scyphistoma) to the sexually active jellyfish (medusa). This not only is a dramatic reorganisation of the body plan, it may also be accompanied by the emergence of new medusa-specific cell types and potentially the disappearance of polyp-specific cell types. While the life cycle of *Aurelia* has been described in various textbooks, we identify a stage of pre-strobilating polyp not previously described. This stage has four pouches filled with multiple folds of inner cell layer epithelia (**Fig. 1A**), representing the first external sign of the beginning of strobilation. Because these four pouches look like a cloverleaf from above, we call this stage the "clover-polyp". In order to unravel the cellular basis for this metagenesis, we generated twelve single cell RNA seq libraries using the 10x Genomics platform: eight from different stages of the life cycle of *Aurelia coerulea* including two replicates of the sessile polyp, a clover-polyp, early and late strobilating polyp, newly liberated ephyra juvenile, fed ephyra, and the newly metamorphosed medusa (1 month, ~1.5cm diameter); four tissue libraries from a growing medusa (~8cm diameter) including the umbrella (excluding the margin), the margin (including tentacles), the manubrium (including the arms and mouth), and the gastrodermis (including future gonad) (**Fig. 1A**). All processed cell suspensions showed a viability estimate of over 75% using a fluorescein/propidium iodide assay for quantification. We generated a custom mapping tool by combining an in-house *de novo* transcriptome assembly with the genomic annotations from the available *Aurelia* genome showing the best alignment ((17): see methods). This was necessary in order to recover mapping of known transcripts that were missing from the genome assembly. The resulting mapping tool contained 51,485 transcripts with a BUSCO completeness of 94.8% and a duplication rate of 25.9% with only 4.3% of metazoans USCOs missing. The resultant transcriptome undoubtedly contains uncollapsed isoforms but nonetheless provides the best recovery of mappable transcripts. Each single cell library was sequenced to at least 70% saturation, from which we detect 43,889 gene models of which 34,301 are represented by 10 or more unique reads across the entire dataset. Further, ~70% of these gene models do not have readily identifiable orthologs and thus represent putative orphan genes.

The transition from the sessile polyp to the free-swimming medusa is marked by an increase in expressed gene repertoire.

We first asked whether we could identify any gene sets specific for each life-cycle stage (polyp: 8633 cells; strobila: 8805 cells; ephyra: 15052 cells; medusa: 15833 cells) by comparing stage-specific samples against one another. To accomplish this we randomly down-sampled

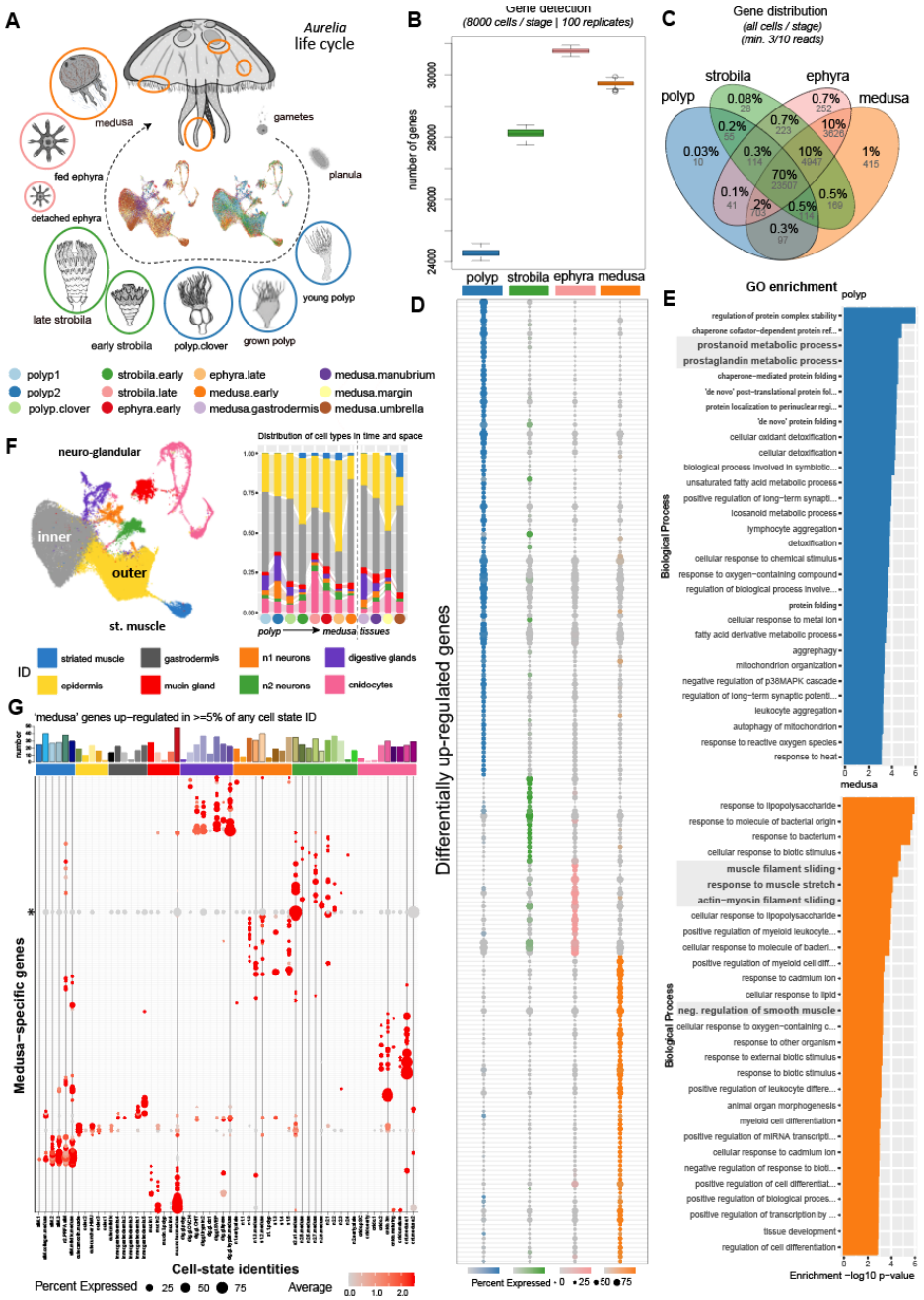


Fig. 1 Single cell transcriptome atlas covering the polyp-medusa life cycle transition demonstrates shifts in genome usage between key life history stages.

(**A**) Twelve single cell transcriptomic libraries were generated from cell dissociations of eight sequential life history stages and four tissue compartments from a sub-adult medusa. Samples are color-coded according to their life-cycle. In the center an UMAP cell plot of the full dataset illustrates the distribution of cells from each library plotted with the polyp samples on top (left) and medusa samples on top (right). (**B**) Distribution of genome use across the life cycle. Number of gene models (>3/10 reads) detected for cells binned according to stage, down sampled to 8000 cells representing each stage, with 100 replicates. (**C**) Venn Diagram illustrates the overlap of detected genes from a single replicate of (B). (**D**) Differentially upregulated genes between life cycle stages, represented as a dotplot. (**E**) Over-represented GO-terms present in the up-regulated gene set highlights prostaglandin synthesis in polyps and overabundance of muscle-related genes in the medusa. (**F**) Distribution of cell type states across all life cycle stages plotted as a 2d UMAP with cells coloured according to their population identity (left) and as a fraction of all cells captured (right). (**G**) Expression of genes not detected in either polyp nor scyphozoa in figure (C), filtered to include only genes up-regulated in at least 5 % of any cell state in the dataset. The number of genes in each cell state is quantified above. Cell partitions are as in (F) and cell states are as in Figure 2. Medusa-specific cell states are highlighted with black outline. * indicates the *aurelin* gene with the highest read count of all medusa-specific genes.

to 8000 cells per stage from the collection of stage-specific samples and then calculated the number of genes detected at each stage. We confirmed this observation with 100 replicates; the number of expressed genes detected was consistently ~20% lower in polyp samples with respect to the other life cycle stages (**Fig. 1B**). We examined the overlap of the detected gene sets and found ~12% of expressed genes are not found in the polyp nor strobila samples (**Fig. 1C**), indicating that a small portion of the coding genome is activated only in the medusa phase of the life cycle. We then queried the identity of the medusa-specific gene sets and found an enrichment in signalling and ion channels (**Fig. S1A, Data S1.1a**), suggesting a possible increase in neuronal diversity within the medusa compared to the sessile polyp stage. The vast majority of these 'medusa' genes are expressed at very low levels (< 50 reads total), and thus could represent false negatives, so we filtered for genes that were up regulated in at least 5% of any cell state identity (as described below). While all medusa-specific cell-type identities express multiple of these 'medusa-specific' genes (**Fig. 1G** upper bar graph), these are also detected across most other clusters in cell-state specific expression profiles (**Fig. 1G** lower dotplot). Nearly half of these genes have no orthology information associated with the gene models, and those that have functional annotations indicate that these genes form part of the cellular effector gene sets (**Data S1.1b**). Given the low detection levels of these genes, we also examined differences in over-abundance of gene expression across the life cycle by calculating the set of differentially expressed genes across all four life cycle stages (**Fig. 1D, Data S1.2**) and found that the polyp gene complement is enriched for prostaglandin synthesis (**Fig. 1E**). Interestingly, it has been demonstrated that blocking prostaglandin signalling with indomethacin can induce strobilation in this species (18). In contrast, the medusa stage, with its large muscular swimming bell, is characterised by an enrichment of collagen and muscle-related genes (**Fig. 1D,E, Data S1.2**). The upregulation of collagen genes reflects the thickening of mesoglea in the medusa, while the muscle-related genes are employed in the expansion of the muscle field in the subumbrella of the medusa. We also note an upregulation of genes involved in mitosis and signalling pathway receptors within the strobila, consistent with the tissue reorganisation that is taking place at this stage (**Fig. 1D, Fig. S2, Data S1.2**).

Transition to free-swimming medusa is accompanied by increased cell-type complement.

We next sought to generate an inventory of the cell diversity present at each point of the life cycle. All twelve datasets were integrated using a reciprocal PCA anchor approach (19) (**Fig. 1A, S1B**). We identified 9 broadly defined cell populations, for which we assign identities by assessing up-regulated gene lists (**Data S1.3**). Principal cell populations include cnidocytes (pink), two major classes of neurons (orange, green), digestive (purple) and mucin (red)

glands, striated muscle (dark blue), and two large populations that represent the inner (grey) and outer (yellow) cell layers (**Fig. 1F**). Of these, cells within the striated muscle cluster are absent from the polyp-derived samples. To obtain a deeper characterization of cell type dynamics in this system, we investigated the substructure of each population in terms of cell-type distribution (**Fig. 2A**) and up-regulated genes and DNA binding molecules (**Data S1.4:12**). We consider all distinct transcriptomic clusters identified here as representative of cell type states, as in some cases identified clusters could represent different cell states of the same cell type. We used the aligned principal components from the integrated dataset to build a cell state tree for all identified clusters (**Fig. 2B**). The cell state tree indicates the majority of transcriptomic identities cluster according to partition and thus can be considered cell type families. Two cell states fall outside of their partition of origin with this approach (highlighted grey in tree, **Fig. 2B**). In total we find thirteen cell-type states that are not represented (<3 cells) in the polyp scRNA sequencing libraries (arrowheads; **Fig. 2B**).

Overall, the relative proportion of cells identified as the inner and outer epithelia remains constant across the life cycle. The polyp stage is characterised by having a larger set of digestive gland cell types (**Fig. 1G, 2A** shades of dark purple), which become localised to the gastric filaments, manubrial arms, or mantle margin in the medusa (**Fig. S3**). There are no digestive gland cells captured from the umbrella sample, and the relative numbers of digestive gland cells captured during the non-feeding transition stages is reduced. There is one digestive cell state that is not found in the polyp samples ("dig.gl.8.medusa"). Digestive gland cell types all share the expression of two *achaete-scuteC* (*ashC)* orthologs (**Fig. 2C:** purples) but have differing protease profiles (**Fig. S3**). All mucin subtypes (**Fig2:** reds) express a *neurogenin* (*ngn)*-ortholog (**Fig.2C**; reds), which is characteristic for neuronal cell populations in other systems (13,20–23). The mucin-producing population also contains one medusa-specific subtype (**Fig. S4**). Additional medusa-specific cell types include the striated muscle and several cnidocyte and neuronal subtypes described below.

We next assayed the differential expression of transcription factor families that are over-represented or under-represented in a cell-specific manner compared to their contribution to the genome. Of note, bHLH (basic helix-loop-helix) protein families, homeodomain (HD) genes, forkhead (FOX) transcription factors, and the bZIP protein family are over-represented as terminal cell type markers, while the number of zinc-finger proteins of the N2C2 class are under-represented (**Fig. 2D, Data S1.4b**). This suggests that determination of the final cell phenotype is governed by specific classes of transcription factor families in a non-random way.

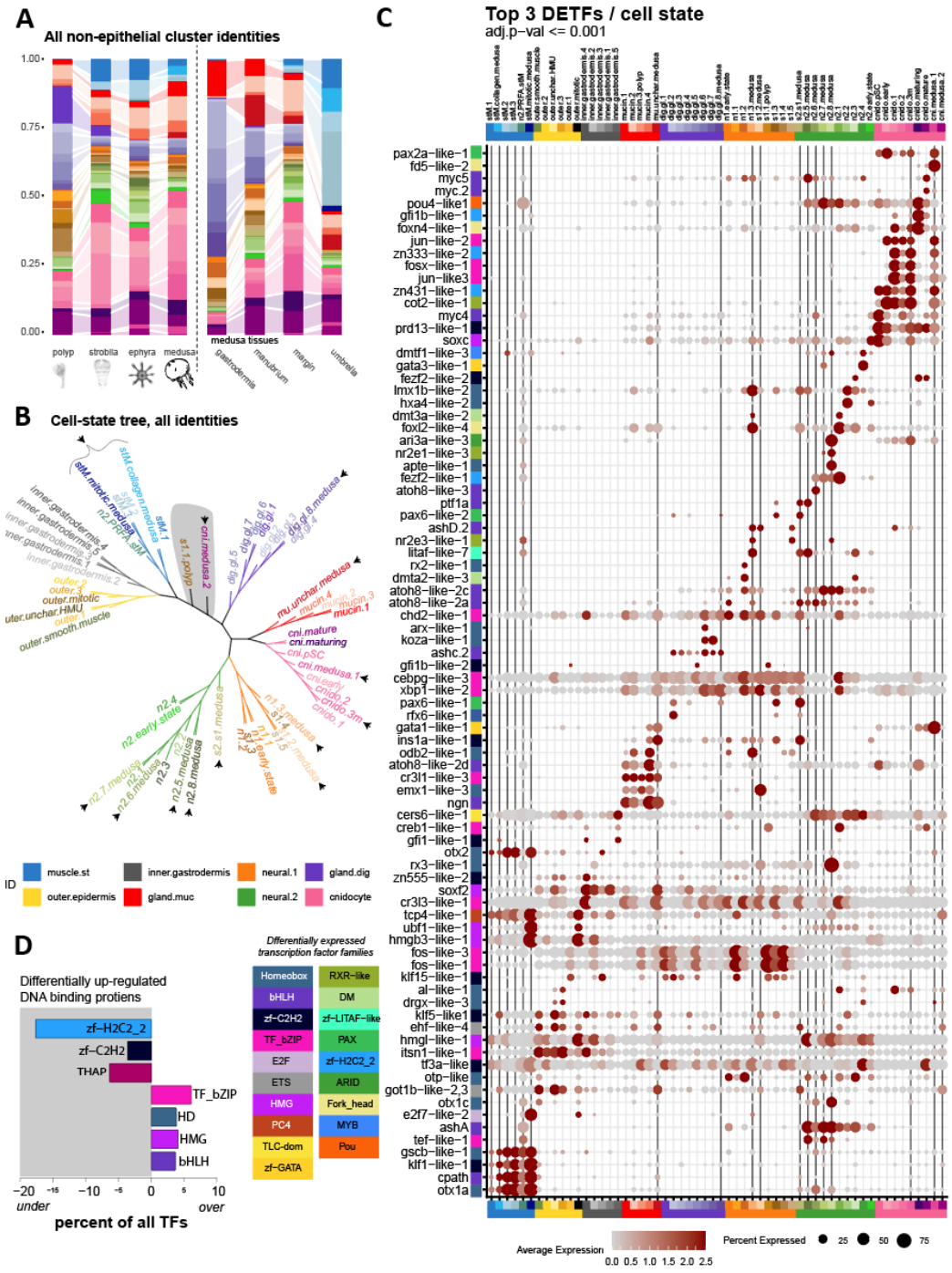


Fig. 2 Integrated cell atlas permits identification of cell states across the life cycle. (**A**) Percent of identified cell type-states across each life history stage and medusa tissue. Epithelial sub-states are not included. Colour code as in (B) and (C) (**B**) Cell state tree illustrating the principal cell-type families present in the dataset. Medusa-specific cell states are highlighted with arrowheads. (**C**) Top five differentially expressed transcription factors across all cell states in the atlas. All transcription factors are colour-coded according to type, colours as in (D). Medusa-specific cell states are highlighted with black gridlines. Full gene lists are available in **Data S1.4b**. (**D**) Transcription factor families that are over or under-represented amongst up-regulated genes compared to their distribution in the genome; only families that differ by at least 2% of all expressed genes with DNA-binding domains are illustrated.

Cnidocytes highlight conserved features of cnidogenesis between anthozoans and scyphozoans

Recently we reported that within the sea anemone *Nematostella vectensis*, specification of the distinct cnidocyte types is marked by a diverging transcriptomic profile corresponding to the formation of the different capsule types, which then switch to converge upon a neural-like expression profile, demarcated by up-regulation of *GFI1B* (13). Notably, we find a similar forked trajectory within the cnidocyte population of *Aurelia*. (**Fig. 3A**). There is a cluster of 'early' cells that separate along three principle trajectories (*cnido.1*, *cnido.2*, and *cnido.3m*), which then converge upon a second mature transcriptomic phenotype. One of the specification trajectories (*cnido.3m*) is absent from the polyp and medusa-gastrodermis samples (**Fig. 3B**). There are two additional unique transcriptomic states within this partition, which are also absent from the polyp (**Fig. 3B**: *medusa.1* and *medusa.2*). One of these cell states expresses the gene coding for the antimicrobial peptide Aurelin (*aure*), in addition to many genes characteristic of cell-types from other partitions (**Fig S5**). This could represent a putative progenitor cell state, be indicative of cell multiplettes or tightly adhered cell fragments, or even an induced immune response in a cnidocyte lineage. Further experiments are necessary to distinguish between these possibilities. The other medusa-specific cell cluster, "*medusa.1*", expresses all of the cnidocyte-specific regulatory molecules (**Fig. 2C**), and a unique set of endopeptidases (**Data S1.9**), indicating this may be a medusa-specific cnidocyte-related cell type localised to the margin of the medusa umbrella.

We combined the *Aurelia* cnidocyte subset with that from the updated cell atlas for *Nematostella* (14), using the set of identified 1:1 orthologs as determined by OMA (24). We first removed cell clusters annotated as being 'planula' or 'gastrula' from the *Nematostella* dataset (**Fig. 3D**). We then aligned the principal components across all samples from both species using the harmony algorithm (25) and used these components to generate a cell state tree (**Fig. 3C**). Both early and late phase cnidocyte cell states from both species cluster together, whereas amongst the sub-type specification pathways we find support for *Aurelia* sub-type *cnido.1* and *Nematostella* sub-type *nem.1* with moderate support. The remaining *Nematostella* sub-types cluster together to the exclusion of the *Aurelia* subtypes. These data suggest that *Nv.nem1* and *Ac.cnido1* represent the isorhiza, which are proposed to be the only cnidocyte type shared between anthozoans and medusazoans (26). We identify a temporal sequence of transcription factor expression that follows the specification trajectories of all cnidocyte subtypes (**Fig. 3E**). We find the earliest cell states express a myc ortholog (*myc4*), a high mobility group (HMG) protein (*soxC*), a prdm13 ortholog (*prdm13a*), and a C2H2-type zinc finger protein (*ZN431-like-1*) similar to *Nematostella* (**Fig. 3E: upper**). The

end of the separate specification trajectories in both species is marked by *paxA* expression and activation of JUN/FOS orthologs, followed by expression of a *gfi1b* zinc finger ortholog marking the switch to a more neural-like transcriptomic profile and activation of *soxA* and additional *myc* orthologs (**Fig. 3E: lower**). One core difference between the species is the late activation of *myc3* in *Aurelia* compared to *Nematostella*, whereas all other *myc* orthologs have similar temporal expression profiles within the cnidocyte lineages (**Fig. 3F**).

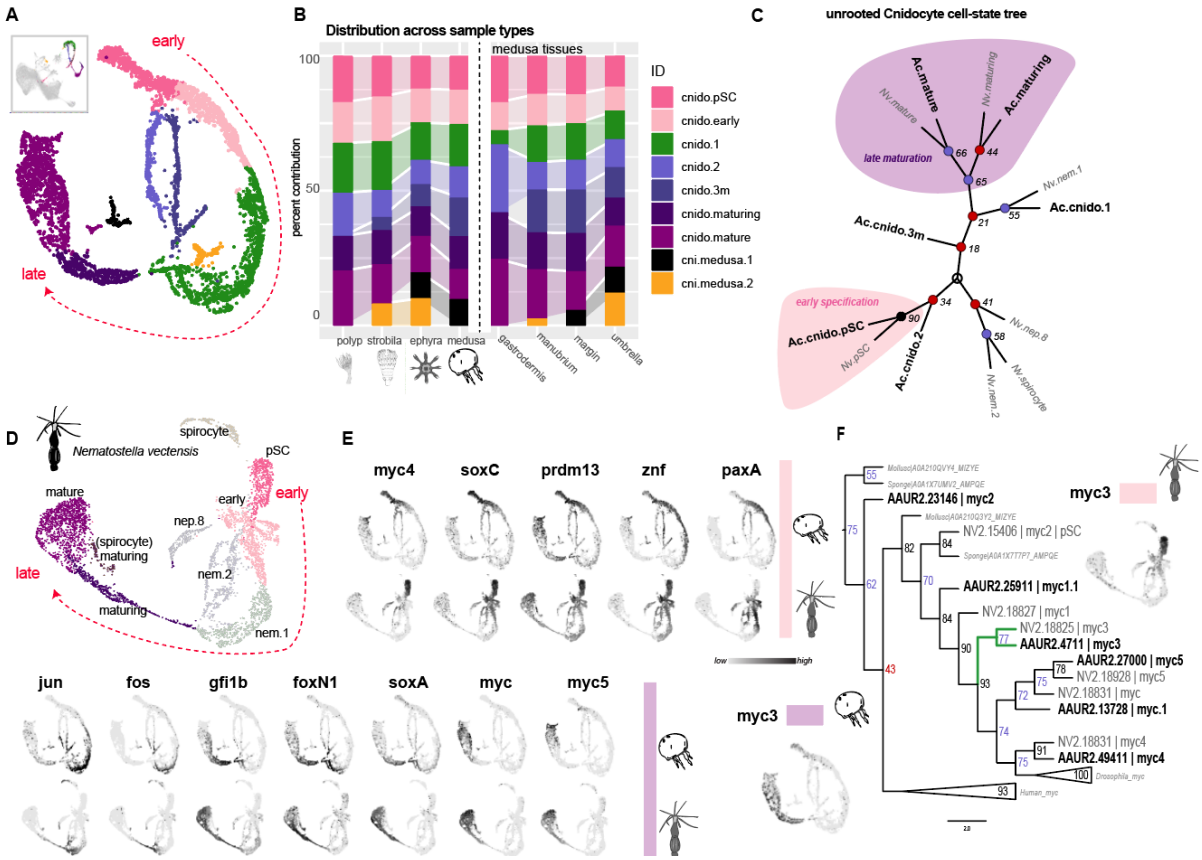


Fig. 3. Cnidocyte specification pathways indicate two major capsule types. (**A**) Cnidocyte trajectory (UMAP) and (**B**) distribution (bar chart). Inset: location of subset within the full dataset. (**C**) Neighbour-joining cell state tree of cnidocyte profiles from *Aurelia* and *Nematostella* further indicate putative homologies between early (pink highlight) and late (purple highlight) phases of cnidogenesis. (**D**) Cnidocyte complement in *Nematostella vectensis* adapted from (14). The planula-specific cell clusters are removed. (**E**) Expression profiles of core transcription factor orthologs in *Aurelia* (top) and *Nematostella* (bottom) indicative of both early specification (pink) and late maturation (purple). (**F**) Gene tree of *myc* orthologs from both species. Paralogs show similar expression patterns in both species except for myc3 (green), expressed early in *Nematostella* (top right) and late in *Aurelia* (bottom left). Bootstrap support values in (C) and (F) are indicated at the nodes; black indicates >80%, purple <80>50%, and red nodes are supported by <50% bootstrap.

Aurelia neural complement reveals two neural classes with similarities to anthozoan neurons.

The anthozoan *Nematostella vectensis* has two principal neural sub-families that have been described that correspond to those with *insulinoma* expression (n1) and those with *pou4* expression (n2) (13,14). Similarly, within the *Aurelia* single cell data we find two distinct transcriptomic partitions with neural properties ("neural.1" and "neural.2") classified here as class 1 and class 2 subtypes (**Fig. 1A**). We investigated the distribution of cell states within both neural partitions to generate an inventory of putative neuronal sub-types (**Fig. 4A,B**). Similar to the distribution described in *N. vectensis*, *A. coerulea* class 1 neurons (orange) are *insulinoma* (*ins1*) positive and class 2 neurons (green) are *pou4* positive (**Fig. 4C**). Class n1 includes 3 neuronal sub-types: "*n1.1*" and "*n1.3*" that express acetylcholine receptors, and "*n1.2*" that is enriched in LWamide expression (**Data S1.10**). The class n1 family also includes putatively non-neural secretory cell types ("*s*"), which are enriched in genes associated with digestion and extracellular matrix production (**Data S1.10**). These data suggest a close relationship between neurons and gland cells, like what has been suggested in other cnidarians (13,27). We find one 's' class cell type originally clustered within the class 2 neuron partition ("*s2.s1.medusa*"), which expresses high levels of the antimicrobial peptide Aurelin (*aure* **Fig. S5**) and is also *ins1* positive (**Fig. 4C**). This '*s*' class cell type appears first in ephyra and is found predominantly in the gastrodermis and manubrium samples from the medusa and is absent from the umbrella (**Fig. 4B**). The gene coding for the Aurelin protein is also the medusa-specific gene with the highest expression and is detectable across most cell states ("*" in **Fig. 1G**). In *Aurelia*, most neurons in the polyp stage are class 1 (**Fig4B:** shades of orange - ~30% of all neural cells), with over 50% of polyp cells in the neural.1 and neural.2 partitions being of the s1 secretory variety (**Fig4B:** shades of brown). These putative secretory cells are found primarily in the gastrodermis and manubrium samples of the medusa. Class 1 neurons in the medusa are also most prevalent within the gastrodermis and manubrium, and includes one subtype that first appears in the strobila and is found in all medusa tissue samples ("*n1.3.medusa*"; lower black box **Fig. 4F**). Thus, similar to that described for the anthozoan *Nematostella vectensis* (13,14), Class 1 neurons and related secretory cells comprise the predominant type of neuroglandular cells in the polyp stage. Further, these are the primary neuroglandular cells within the gastrodermis of the medusa.

In contrast, neurons of class 2 comprise most neuron types within the umbrella and margin medusa samples and include four neuron clusters that first appear at strobilation (**Fig. 4B**). Within the polyp, only the poly-RFamide class 2 neurons are abundant ("*n2.2*"), and all other class 2 neural types are under-represented or absent (**Fig. 4B**). *Ins1*-negative class 2 neurons

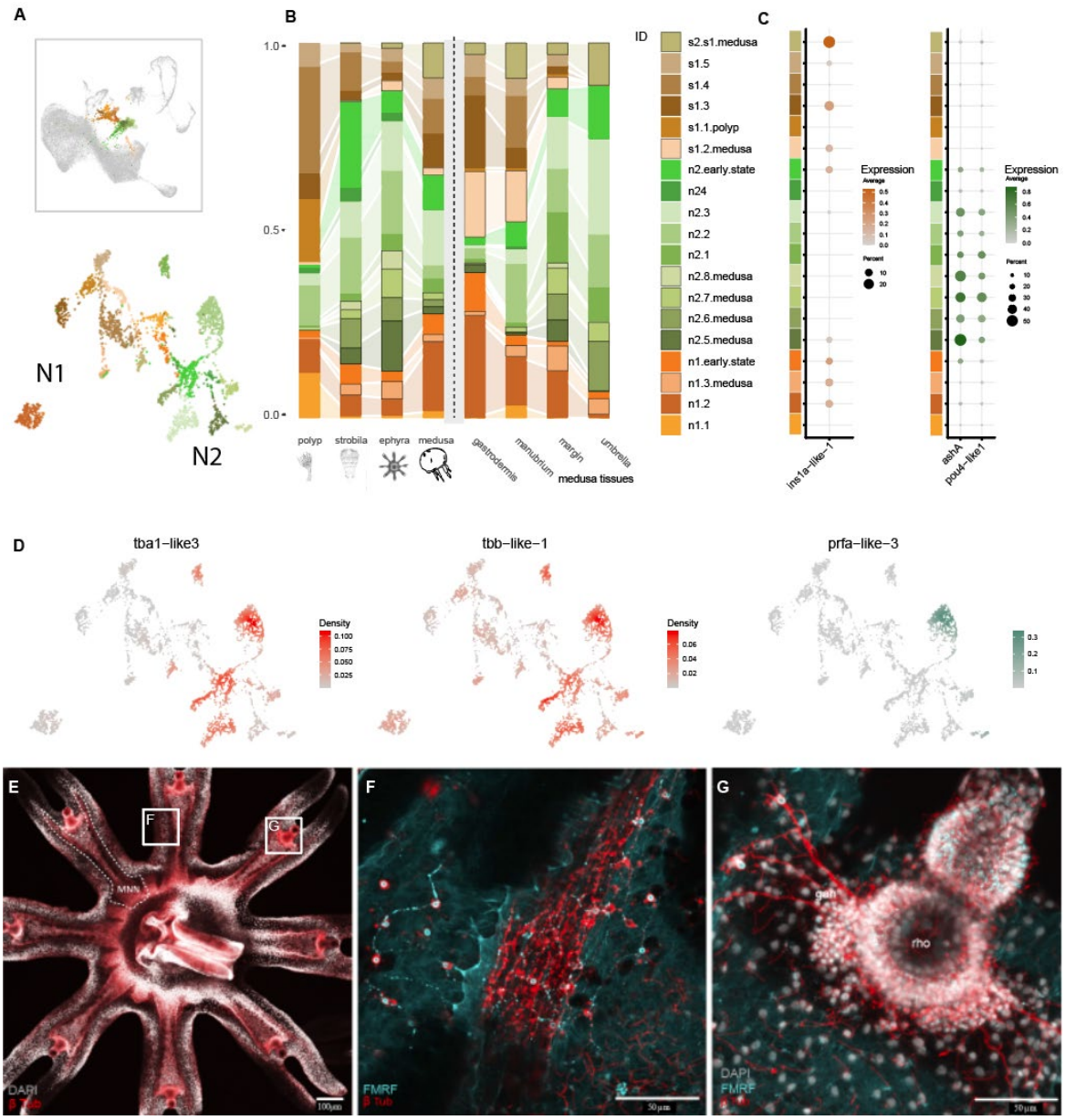


Fig. 4 Neural cell complement includes two distinct classes, with medusa-specific subtypes in each. (**A**) Distribution and characterization of neural complement, illustrated on an UMAP embedding of the full data (top) and neural subset (bottom) (**B**) Distribution histogram across all samples, medusa-specific cell clusters are highlighted with black outline. (**C**) Two classes can be recognized by differential expression of transcription factors. Class 1 cells express *insulinoma* (*ins1a*) (orange left) and class 2 express *achaete-scute* (*ashA*) and pou4 (*pou4-like1*) (green right). (**D**) Density expression plots illustrating N2-specific α-Tubulin (*tba1-like3*), β-Tubulin (*tbb-like1*), and precursorFMRFamide (*pRFa-like3*) expression. (**E**) An anti-βTubulin antibody stains the motor nerve net of the ephyra. (**F**) Co-staining of anti β-Tubulin and anti FMRFamide visualises cells that are positive for both antibodies, possibly corresponding to cells expressing both tubulin and FMRFamide. The ephyra lappet displays condensed regions of the β-Tubulin positive motor nerve net aligning with the radial canal of the lappet. (**G**) Network of the rhopalia nervous system positive for β-Tubulin. DNN = diffuse nerve net, gan = ganglion, lap = lappet, man = manubrium, MNN = motor nerve net, rho = rhopalia.

express both the bHLH transcription factor *ashA* (**Fig. 2C, 4C**), and a *pou4/brn3* ortholog that is shared also with the cnidocytes (**Fig2C**). This is like the situation described for *Nematostella* (13,14). Class 2 neurons (**Fig4B,** shades of green) increase in abundance during the transition to the medusa form, including the generation of four additional medusa-specific class 2 neurons. Three are present already in the early strobila ("*n2.5.medusa*", "*n2.6.medusa*", and "*n2.7.medusa*"), whereas "*n2.8.medusa*" appears only in the late strobila sample. We presume that these cell identities correspond to sensory cells of the rhopalium and putative motor neurons that innervate the striated swimming musculature, as these tissues are medusa-specific and first develop within the strobila. Published expression data further supports this postulation: *Pit1* has previously been reported to be expressed in putative sensory neurons of the developing rhopalia (10), and is expressed in "*n2.5.medusa*". Neuron "*n2.8.medusa*" expresses the homeodomain transcription factor *otx1c* and apterous-like LIM-homeodomain *apte-like*, as well as poly-RFamides (PRFA) (**Fig. 2C**). *Otx1c* has been reported to localise to cells of the rhopalia (28) and hence these cells could correspond to a set of FMRFamide positive sensory cells in the oral-distal portion of the ephyra lappets (**Fig. 4E,G**). The candidate specification marker for neuron "*n2.6.medusa*" is an atonal-like bHLH family transcription factor, *atoh8-like3*. We detect an expansion of the ATOH8 gene family, with detectable expression for 6 gene models. Two are expressed in all class 2 neurons, while others show expression also in cnidocytes and mucin gland (**Fig.S6**). Only *atoh8-like3* ("*n2.6.medusa*" neuron) and *atoh8-like-1* ("*mucin.medusa*") show specific expression restricted to medusa-specific cells. The ortholog of *atoh8-like3* in *Nematostella* is expressed in the two gastrodermal N2 class neurons ("*N2.g1*" and "*N2.5*";(14)), whereas *atoh8-like-1* appears to be an *Aurelia atoh8* paralog. This suggests the "*mu.medusa*" cell type may be a novel cell type that arose after gene duplication, as has been proposed for anthozoan muscle (29), whereas "*n2.6.medusa*" may be a cell type ortholog of the n2 gastrodermal neurons in the anthozoans. This is an intriguing observation as the gastrodermal neurons lie in close proximity to the musculature of the inner cell layer (30).

We next sought to connect the single cell transcriptomic data to the anatomy of the nervous system. The scyphozoan nervous system is composed of three distinct parts: a diffuse nerve net (DNN) that covers the entire animal and is comprised of four anatomically distinct neural structures, the giant-fiber or motor nerve net (MNN) that is comprised of larger bipolar neurons that overlie the muscle fields, and the marginal ganglia or rhopalia that is located at the tips of the ephyra and later the margin of the medusa (31). We find that class 2 neurons all express elevated levels of specific alpha- and beta- tubulins (*TBA1-like3* and *TBB-like-1*; **Fig. 4D**). Anti-β-tubulin antibody staining highlights concentrations of the MNN centrally located within the lappets of the ephyra ((32) **Fig. 4A**). Morphologically, this is the territory that includes the radial

canals and overlying muscle fields of the ephyra. FMRFamide antibodies stain a subset of β-Tubulin positive portions of the medusa nerve net (**Fig. 4B**). Whereas the β-Tubulin antibody recognizes cells predominantly found near the striated muscle field and ganglion cells projecting out of the rhopalia, FMRFamide positive neurons are more scattered across the ectoderm (**Fig. 4C**). Some FMRFamide positive neurons are also found within the distal-most portion of the rhopalia (**Fig. 4C**). This distribution suggests that the β-Tubulin antibody targets the MNN, whereas the FMRFamide positive neurons form part of the DNN as previously reported (12). Specific tubulin-paralog expression within the class n2 neurons suggest that this is the portion of the nervous system labelled by the β-Tubulin antibody.

Striated and smooth muscles share partial transcriptomic profiles.

The striated musculature is one of the specific features of the medusa stage. Striated muscles have in general not been described within the polyp except for one study that had postulated areas of striation in myofibers at the base of polyp tentacles (33). We investigated the profile and origin of the striated muscles in Aurelia. A clear upregulation of muscle-related structural genes such as myosin, actin, and actin-binding proteins is indicative of contractile properties, and thus expected to be of either smooth or striated muscle nature. One of the main partitions in the single cell transcriptome dataset defined by a large set of muscle-related genes appears at the strobila stage (**Data S1.3**). This cluster contains a large cellular contribution from the umbrella sample and is missing from the manubrium and gastrodermis tissue libraries (**Fig2A**: dark blues). To our knowledge the later tissues contain only smooth muscle. This indicates that this cell cluster is the striated musculature. In the sea anemone *Nematostella*, anemone-specific gene duplications of members of the PaTH (Paraxis, Twist Hand-related) bHLH family of protein coding genes was driving the diversification of muscle cell types (29). Interestingly, the single-copy ortholog of the PaTH gene, called *cpath* (29) is expressed in the putative striated muscle cluster. Surprisingly, otx paralogs, *otx1a*, [aka Otx1 (10)]**,** and *otx2* are upregulated in the striated muscle cluster (**Fig. 2C**). *Otx1a* expression was previously shown to overlap with the striated muscle fields, however was inferred to define the overlying neuroectoderm of the motor nerve network (10). The current data does not exclude neural expression, as this gene is also detected within some neural populations albeit to a lesser extent (**Fig. 2C**). We identify an additional sub-population of cells with a clear muscle signature (eg. *myophilin/calponin*, *myosin regulatory and essential light chains, myosin heavy chain*: **Data S1.5**) within the outer epithelia population, which we annotate as the smooth muscles. The identity of these two muscle cell populations is confirmed with *in situ* hybridization (**Figs. 5**-**7**).

To investigate the dynamics of muscle formation, we further compared phalloidin staining of muscle fields with *in situ* hybridization detection of specific cluster marker expression in polyps (**Fig. 5**), strobila (**Fig. 6**), and ephyra (**Fig.7**). We first assessed the anatomic location of the muscle fields (**Fig. 5A,B**) with phalloidin staining of actin filaments in *Aurelia* polyps (**Fig. 5C-H**) and ephyrae (**Fig. 7C-F**) corroborating previous studies using electron microscopy (33,34). Polyp muscles form three distinct muscle fields (**Fig. 5D-H**): the radial muscles of the oral disc (**Fig. 5E**), the longitudinal tentacle muscles (**Fig. 5F**), and the longitudinal retractor muscles that run along the body column (**Fig. 5G,H** (35)). After detachment of the entire stack of ephyrae, the remaining aboral part of the polyp (**Fig. 6A,B**) regenerates into a functional polyp (**Fig. 6G-I**). In the freshly detached ephyra, two striated muscle fields are present (**Fig. 7B-D**) (34). One muscle field consists of radial striated muscles that extend into the ephyra lappets (**Fig. 7C**), the other is a field of circumferential striated muscles that surround the oral disc (**Fig. 7D,E**). Fragments of the polyp smooth retractor muscles remain detectable in early ephyra (8), detectable by *in situ* hybridization of the muscle marker *stmyhc1 (striated myosin heavy chain 1)* (**Fig. 6L-N, P-S**). As medusa development progresses, the radial and circumferential muscle fields merge into a unified umbrella muscle (35). At the periphery of the ephyra striated muscle fields, at the tips of the lappets, we find alternating smooth and striated muscle fibres (**Fig.7F**). These can be observed in both freshly detached ephyra as well as in metaephyrae, an intermediate stage between ephyra and juvenile medusa (36). Further, we find mixed smooth and striated muscle fibre composition at the base of the developing manubrium in metaephyrae (**Fig. 7G-I**). In some cases (**Fig. 7F,H**) zones of striation can be observed within otherwise smooth muscle fibres (**Fig. 7H**).We also confirmed the identity of smooth and striated muscle cell types detected in the scRNAseq data by in situ hybridization (ISH) (**Figs. 5,6,7**). *myophilin-like2* is restricted to smooth retractor muscles of polyps (**Fig. 5L**) and expression is still visible in fragmented retractor muscles in the late strobila, consistent with the single-cell expression profile (**Fig. 6D**). *mrlc2* and *stmyhc1* reads are detected in both muscle types (**Fig. 7pFig. 5M, Fig 6C,E,G-P, Fig. 7J-L,N-P**), and ISH indicates that the expression is localised to the fields of striated muscles in ephyrae (**Fig. 7J,K,N**), as well as the smooth muscle populations in polyps including longitudinal tentacle muscles, radial muscles of oral disc and retractor muscles of the body column (**Fig. 5M, Fig. 6H,I,L,M**), and the muscles of the manubrium in the meta-ephyra (**Fig. 7L,O**).

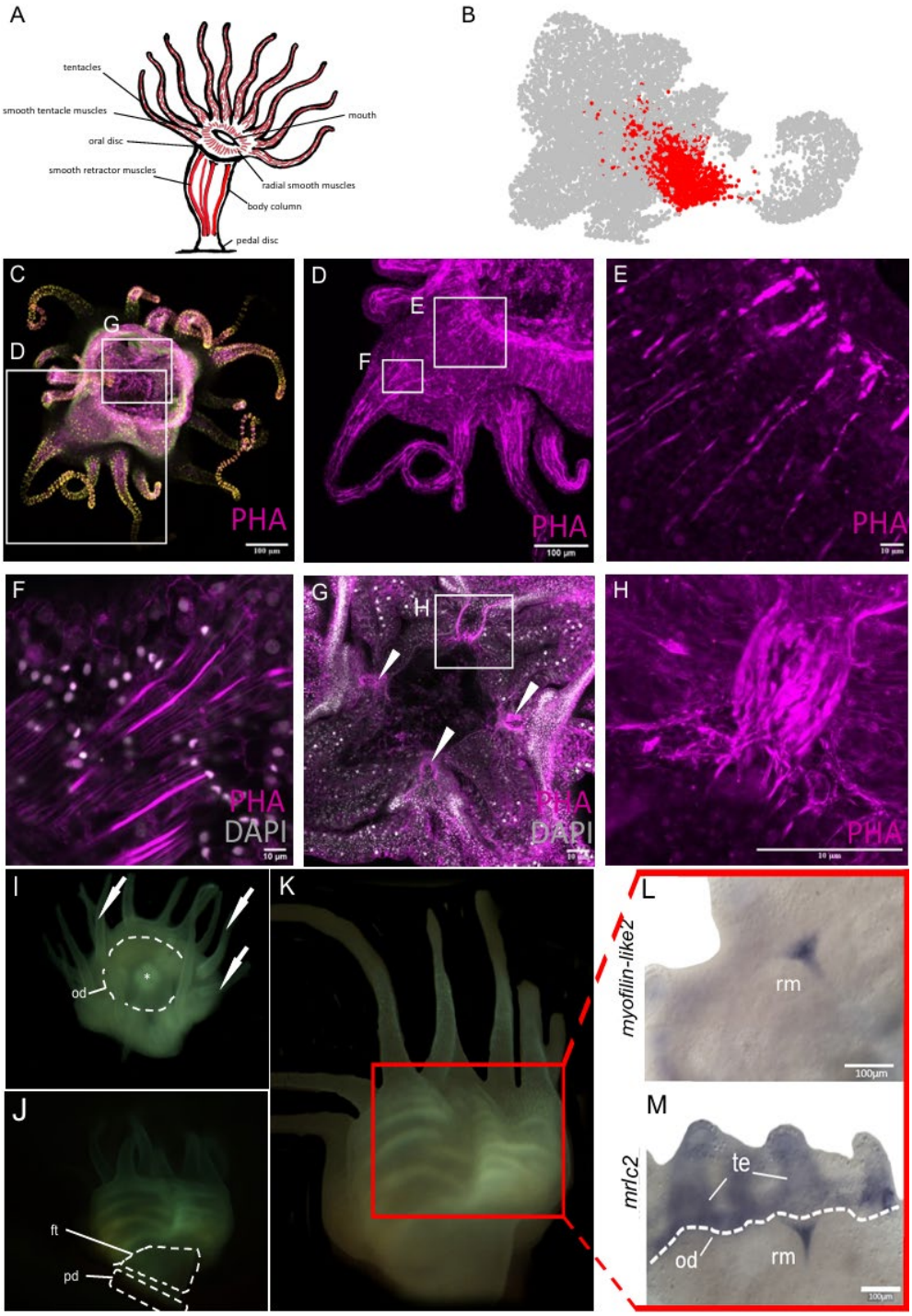


Fig.5. Muscle territories in the *Aurelia* polyp. (**A**) Schematic representation of the different smooth muscle territories in the polyp from a lateral view. (**B**) A reference UMAP of the muscle subset with a highlighted portion of smooth (red) muscle cells. (**C-H**) Phalloidin stains actin-rich muscle fields in the animals (arrowheads point to retractor muscles). (**I-K**) Polyp images illustrating polyp morphology (left, arrows point to tentacles, od = oral disc, asterisk = mouth, ft = foot, pd = pedal disc) and the body column (right, labelled with a red square) to show the location for the gene expression in L-M. (**L-M**) In situ hybridization for two marker genes (*mrlc2; myophilin-like2*) confirms the spatial location of the retractor muscles, tentacle muscles, and muscles at the oral disc of the polyp. te = tentacle, od = oral disc, rm = retractor muscle.

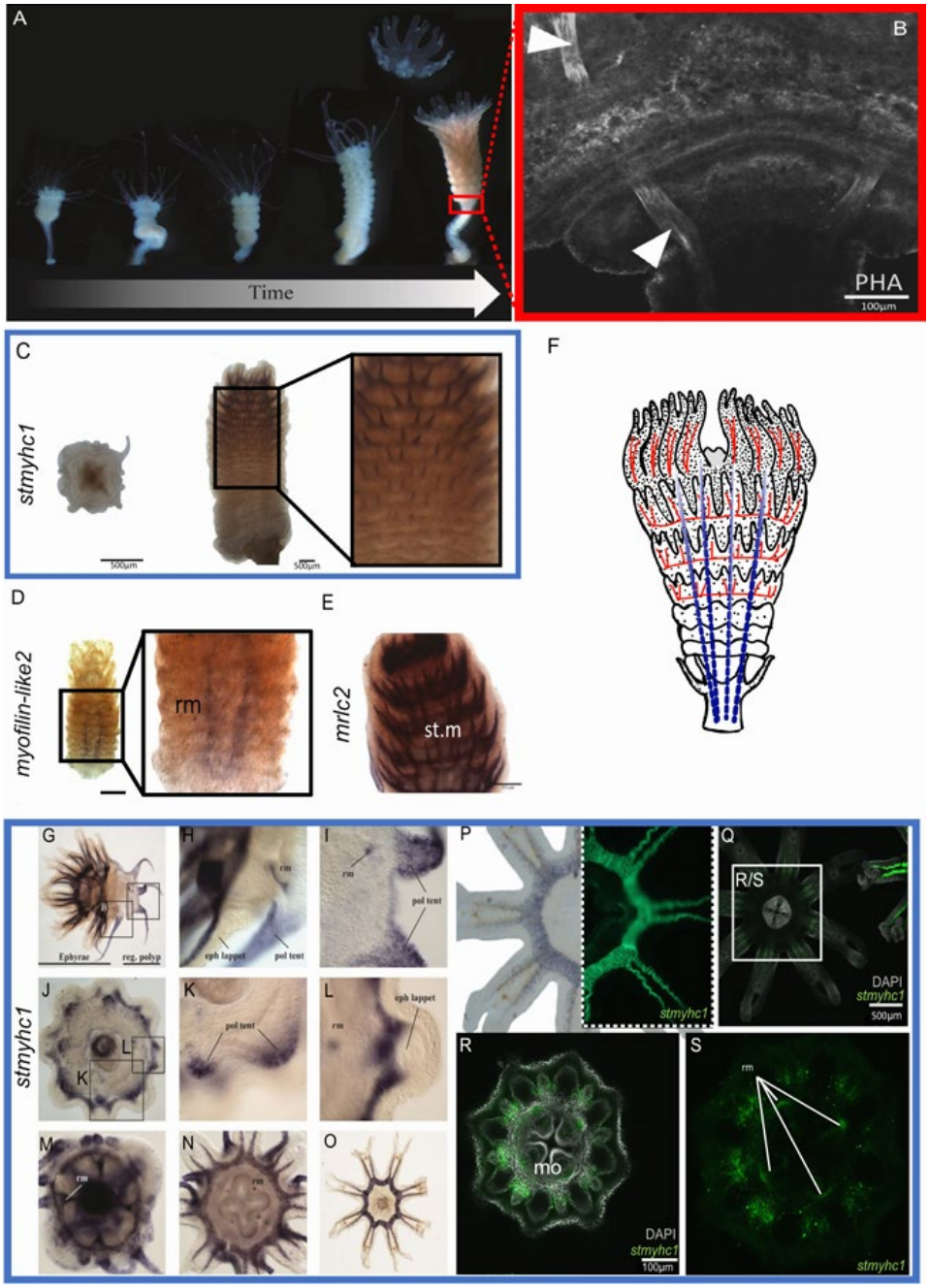


Fig. 6. Establishment of striated muscle field during strobilation. (**A**) Consecutive stages of strobilation from polyp (scyphistoma) to strobila and ephyra. (**B**) Foot region of the strobila stained with phalloidin showing polyp retractor muscles (arrowheads). (**C,D,E**) ISH for marker genes that label either smooth or striated muscles in the strobila. (**F**) Schematic for advanced strobila with labelled muscle fields (red) as well as fragmented retractor muscles (blue). (**G-O**) gene expression patterns of *stmyhc1* in the advanced strobila. (**P-S**) gene expression pattern for the same gene in the detached ephyra. st.m = striated muscles, rm = retractor muscles, mo = ,mouth.

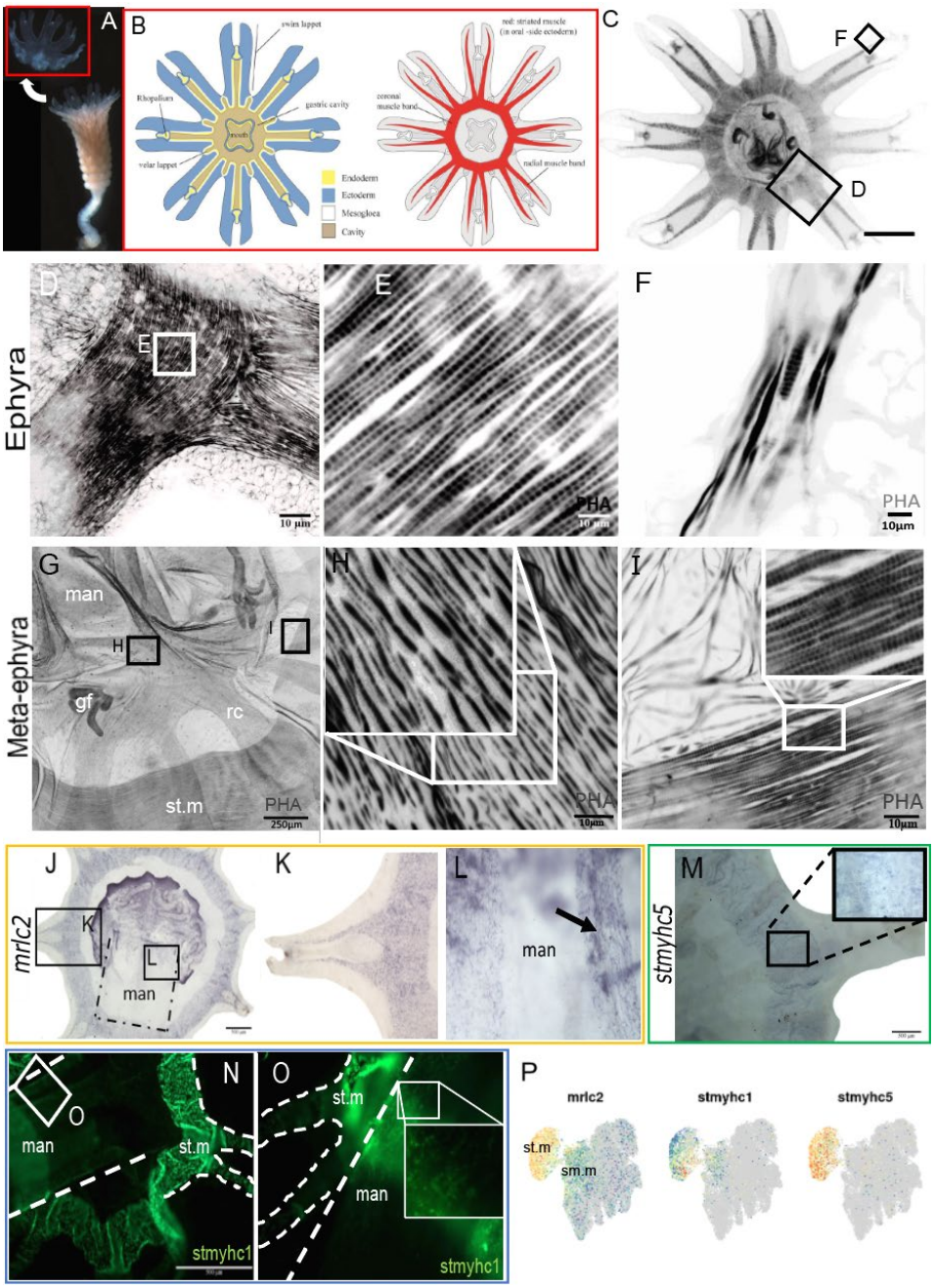


Fig.7: Muscle territories in the ephyra. (**A**) Ephyra detachment from the strobila. (**B**) Schematic of the ephyra body plan illustrating tissue anatomy (left) and striated muscle field (right). (**C-I**) Phalloidin labels the different muscle fields in the ephyra that are composed of either smooth, striated or a mixture of both muscle types. (**J-O**) In situ hybridization for *striated type Myosin Heavy Chain5* (*stmyhc5*), as well as genes shared by both muscle types (*stmyhc1, mrlc2*) identified from single-cell RNAseq data. (**P**) An expression plot of striated muscle specific genes (*stmyhc5*) as well as genes that share their expression profile between smooth and striated muscles (*mrlc2, stmyhc1*). man = manubrium, gf = gastric filaments, rc = radial canal, st.m = striated muscle, boxes color code: red = (A) detached ephyra and (B) corresponding scheme of ephyra morphology; yellow (J,K, P); green (M, P); blue (N,O,P); gene expression profile and corresponding expression from single-cell data shown in a dotplot

Aurelia muscles show a clear molecular signature specifically enriched in muscle related structural genes. We generated lists of muscle-type specific variable genes that are up-regulated in at least 10% of cells in either muscle cluster, and detectable in less than 40% of any non-muscle cluster. While smooth and striated muscles are clearly distinguishable by the differences in their molecular content, a substantial fraction (~40%) of both muscle profiles consists of genes with unknown affinities (**Data S1.13**). Notably both muscle subtypes express a set of genes that comprise a core contractile complex that includes *calmodulin*, *myosin light chain kinase (mylk)* and *tropomyosin* (*tpm*) together with myosin heavy, light, and regulatory chains (*myhc, mle, mrlc*) (**Fig. 8A**; **Data S13**) that is shared with bilaterian smooth muscles (37,38). However, in most cases each muscle subtype both share expression of a common ortholog (**Fig. 8A**) and additionally express a different set of paralogs across the two muscle types (**Fig. 8A**). An exception is the *myosin regulatory light chain* (*mrlc*) and its putative interacting protein (*mylia*), of which there appears to be a single ortholog expressed in both muscle types. Taken all together these data suggest that the contractile apparatus in the Scyphozoa, using here *Aurelia* as a proxy, is similar to the bilaterian smooth muscle contractile complex (**Fig. 8C**).

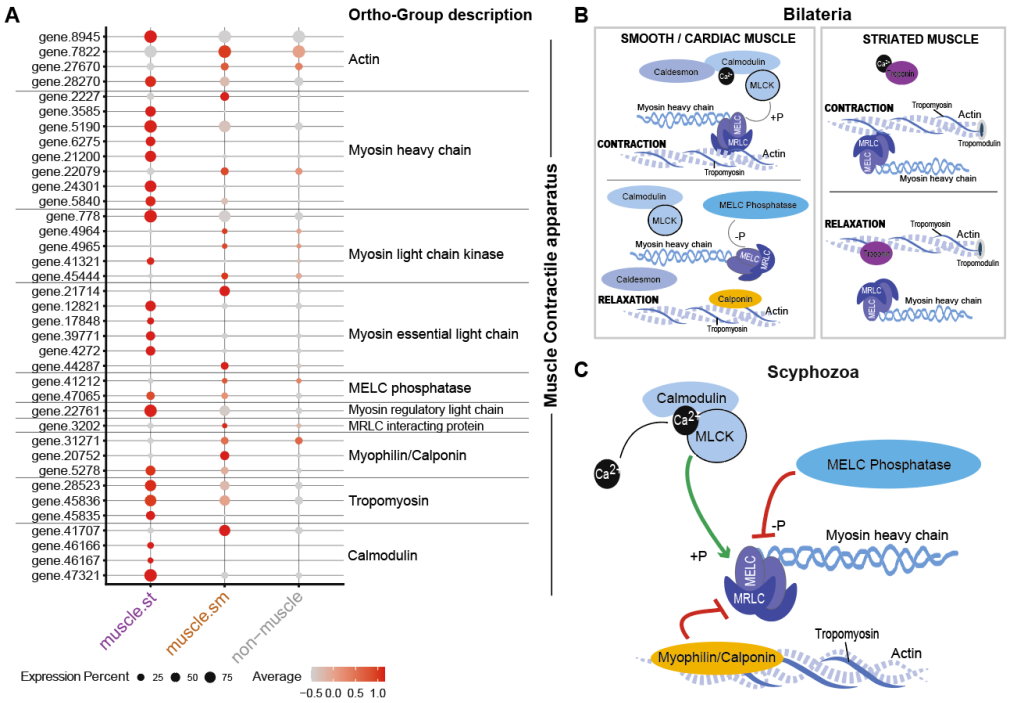


Fig.8: Shared contractile complex between smooth and striated muscle cells. (**A**) Dotplot of gene expression for muscle-type specific gene sets. (**B**) Schematic shows main molecular components for myosin-based contraction in bilaterians. (**C**) Hypothesis of muscle contractile apparatus in Scyphozoa.

Discussion

Single cell transcriptomic atlas reveals distinct cellular subtypes associated with the formation of medusa in the moon jelly.

The life cycle of the sea anemone *Nematostella vectensis* has been characterised at a single cell level by recent studies (13,14,39). However, the medusozoans typically have a more complex, metagenic life cycle with a transition of the asexually reproducing polyp to the sexually active medusa. The formation of these two distinct adult life forms has fascinated researchers for more than a century (6,28,40,41). Within medusozoans, hydrozoans, scyphozoans and cubozoans form medusae in dramatically different ways, raising the question whether they are homologous or convergent (28,42). Recent single cell transcriptome studies in hydrozoans have analysed the freshwater polyp *Hydra* (27), which has lost the medusa stage, the medusa stage of *Clytia hemisphaerica* (43), and a series of medusa, 'four leaf structures', 'cysts' and polyps characterizing reverse development in *Turritopsis rubra* (44). The latter study also compares *T. rubra* with the medusa of *Aurelia coerulea* collected from the waters near Yantai China in a study of swimming behaviour (44). The molecular changes underlying the transition from polyp to medusa have been characterised via bulk transcriptome sequencing (17,45,46), however changes at the cell type level during the metagenic transition are unclear from these studies. Here we present a single cell characterisation of the metagenesis in a scyphozoan, the moon jelly *Aurelia coerulea* (formerly sp.1), from a strain of unknown origins that has been kept in the laboratory for over 25 years. We document an overall increase in gene detection of approximately 20% during formation of the medusa. Other studies have looked for differential gene use from bulk RNA extracts from medusa versus polyp stages and have found no evidence for increased taxon-specific genes in either stage (17), or even composition of reconstructed proteome or transcriptomes from polyps compared to medusa (47). While initially this would seem to contrast our finding of an overall increase in genome usage in the medusa compared to the polyp, these results are not inconsistent. While we find only one polyp-specific digestive cell type, we document several distinct cell types associated with the generation of the medusa stage that are detectable already in the strobila but absent from all the polyp samples. Apart from the striated muscle, which is an abundant cell type of the medusa, these medusa-specific cell types comprise a small fraction of the overall tissue composition and share transcriptomic similarity with related cell types that are present across the life cycle. Comparisons of differential gene expression clearly demonstrate changes in the overall gene representation across the life cycle (46), in agreement with the current results.

Of the thirteen cell clusters that are absent from the polyp stage, five of these are neurons. We can tentatively associate these neural subtypes with components of the nervous system known to arise in association with the formation of medusa-specific tissues, including the rhopalia, motor nerve net, and striated muscle (10,12,28). We note extensive parallels in the complement of neural subtypes with the organisation described in the sea anemone, in particular the partitioning into two major neural classes. This division is similar to anthozoans, where Steger et al (13) identified insulinoma positive and negative neural trajectories in *Nematostella*. In *Nematostella*, the INS-positive neuron population is more abundant and possesses a greater number of subtypes, including the sensory/secretory "S"- class of largely uncharacterized cell types. It is this N1 subset that also predominates the polyp stage of the lifecycle in *Aurelia*. The digestive gland cell populations in both species express orthologs of the achaete-scute family (*ashC*) (13). *AshC* is an ortholog of the *ascl345* family in vertebrates, which is expressed in the skin (*ASCL4*), salivary glands (*ASCL5*), and teeth (*ASCL6*) [UNIPROT database (48)]. Orthologs of *ashB* and *ashD* are expressed in the neurons in both *Nematostella* (13,14) and *Aurelia,* and both are homologs of the Achaete-scute complex, notable regulators of neurogenesis in bilaterians (49). AshA does not have a mouse or human ortholog, but is present within nematodes, where it is a key regulator of a nociceptive neuron identity (50). In *Aurelia,* the *ashA* ortholog is restricted to the class 2 neurons, while in *Nematostella ashA* expression is found in both types (13,14,51) and *ashD* is restricted to INS-positive (n1) neurons in both species. The INS-negative populations in both species express *pou4* orthologs, found also within the cnidocyte lineages and thus further supports claims of a close relationship between cnidocytes and *insulinoma*-negative/*pou4*-positive n2 neurons (13,14,52). Further, we demonstrate a similar specification trajectory in the formation of cnidocytes as has been described for the sea anemone *Nematostella vectensis* (13,15,53). This strongly suggests that the radiation of achaete-scute genes not only preceded the split of cnidarians and bilaterians, but also that the paralogs were implemented in the diversification of the neuro-glandular lineages across the Metazoa.

Transcription factor family expansions may be related to the generation of novel medusa cell types.

Recently, we demonstrated a sea anemone specific expansion of an orphan bHLH family that correlated with the diversification of muscle cell types (29). Expansions of other transcription factor families in a similar manner could be related to generation of novel cell types (15). The medusa-specific cell types identified here also demonstrate cell-type specific expression of expanded transcription factor families. For example, we find an *Aurelia atoh8* (*atoh8-like3*) paralog expressed specifically in the putative motor neuron ("*n2.6.medusa*"; **Fig. S6**). *Myc*

genes appear to be expanded within the cnidarians, at least in *Nematostella* (6 copies (54)) and scyphozoans (7 or 8 copies; **Fig. 3F**). While four paralogs have been reported from the hydrozoan *Hydra* (55), only 2 have been investigated (56–58). Similarly, only 2 paralogs are reported from another hydrozoan (59). Interestingly, of the two *hydra myc* genes, one is expressed in the cnidocyte lineage, while the other is expressed in the interstitial stem cell lineage (27). The *myc* expansion in *Nematostella* and *Aurelia* are used in parallel ways during cnidocyte development, and orthologous pairs of *Nematostella* and *Aurelia* myc genes are similarly expressed in the early and mature phases of cnidocyte differentiation (**Fig. 3E,F**). In both species, the basal-most *myc* paralog is expressed within the putative stem cells. Within the atonal/neurogenin bHLH TF clade we find two expansions, one shared with *Nematostella* and thus likely representing a cnidarian expansion, and the other specific to *Aurelia* (**Fig. S6, S7**). Within the *Aurelia* branch we find paralogs that define pou4-positive/insm-negative class 2 neurons, but also others that have expression restricted to the medusa-specific putative motor neuron population "*n2.6.medusa"*. These neurons also express paralogs of the expanded *otx* family. Of interest, expression of one of these *otx*-family paralogs, *otx1a*, is shared with the striated muscle, while the *otx2* paralog is exclusively expressed in the striated muscle. These data support the hypothesis connecting gene duplication with cell type expansion (37,60–64), suggesting that at least some of the medusa-specific cell types identified here may represent synapomorphies of the Medusozoa, or perhaps even the Scyphozoa. *Otx* was also found to be expressed in striated muscles of the hydrozoan *Podocoryne carnea* (66). Further, in situ hybridization of *otx* showed expression in the medusa velum of the hydrozoan *Clytia hemisphaerica* (28), which is also characterized by striated muscles. This suggests that its role in striated muscle formation is ancestral to the medusozoans. Of note, in bilaterians, o*tx* genes mark mostly anterior parts of the body, in particular anterior brain regions in a conserved manner. While it has been implicated in repressing muscle development in ascidians (63), to our knowledge has never been implicated in positive regulation of bilaterian muscle development. This supports the view of an independent evolution of striated muscles in cnidarians and bilaterians (37). It further raises the possibility that the striated muscles in jellyfish, like the fast contracting tentacle muscle in *Nematostella* (29) express at least partially neuronal features.

Shared core contractile machinery between smooth and striated muscles indicates a gradual maturation of smooth fibres into the striated phenotype.

Striated muscles enabled jellyfish to evolve a unique swimming behaviour and conquer the oceans. Nevertheless little is understood regarding the origin and evolution of striated muscle cell types. To infer the evolution of this muscle cell type, the molecular identities of striated

muscle from true jellyfish is required. We distinguish muscle cell transcriptomes across different stages of the *Aurelia* life cycle, including both striated and smooth types. The striated muscle population is the largest medusa-specific cell population with a distinct transcriptomic profile. The smooth muscle transcriptomic profile shares features with the outer epithelial layer and is recovered as a cell state within this larger population. While a small set of genes are shared across the two muscle phenotypes (e.g. *stmyhc1* and *mrlc2*), others are more specific to either phenotype (eg. *stmyhc5* in striated muscle; *myophilin-like-2* in smooth muscle) (**Fig. 8A**), which were verified by in situ hybridization (**Figs. 5,6,7**). The shared expression of muscle-related structural genes is most likely part of a core contractile machinery that is present in both smooth and striated muscles (**Fig. 8B,C; Data S1.13**). Interestingly, a functional study done by Tanaka *et al.* (38) linked some of the genes expressed in *Aurelia* striated muscles to a bilaterian smooth muscle contractile apparatus. This core contractile complex is regulated via phosphorylation of *myosin regulatory light chain* (*mrlc*) by *myosin light chain kinase* (*mylk*) in the presence of Ca2+, which is known to regulate bilaterian smooth and non-muscle cell contraction (**Fig. 8B**) (38,65,66). Main players in this smooth contraction mechanism are *calmodulin*, *mrlc*, *melc* and *mylk*, which are all expressed in both *Aurelia* muscle types identified from our single-cell data (**Fig. 8A; Data S1.13**). We found that at least one of the striated myosin heavy chain paralogs (*stmyhc1*) seems to be exclusively expressed in either regenerating smooth or developing striated muscles, but not in mature smooth muscles of polyps (**Fig. 6C,G-O**). This suggests a role for the various myhc paralogs in differentiating muscle architecture. We further find smooth and striated muscle fibres at the periphery of the ephyra striated muscle fields (**Fig. 7F**) and we could find expression of *stmyhc1* at the base of the developing manubrium in the metaephyra (**Fig. 7N,O**)**,** which also showed a mix of both smooth and striated muscle phenotypes (**Fig. 7G-I**). Strikingly, we also observe individual smooth fibres located at the base of the developing manubrium that show striation at some locations along the fibre (**Fig. 7H,I**), raising the possibility that smooth and striated fibres can be found in the same cell. *In situ* hybridization expression data, together with the juxtaposition of both fibre types within the ephyra (**Fig. 7F,G-I**), are suggestive of a gradual maturation of smooth fibres into the striated phenotype. This phenomenon is known from the vertebrate heart muscle, where non-striated pre-myofibrils gradually undergo sarcomere assembly during myofibrillogenesis correlated with the expression of the actin capping gene *tropomodulin* (*tmod*) (67). In scyphozoan species like *Aurelia coerulea* this could be the case during metamorphosis from the polyp to the medusa form. Taking these results together we hypothesise that during strobilation striated muscles could originate from smooth muscle fibers. However, future work is required to further provide experimental evidence for this scenario.

Limitations of the study

Aurelia lab strains have been used as part of developmental, evolutionary as well as ecological studies in the last decades. However, this cosmopolitan species exhibits a lot of cryptic speciation and there are multiple genomic assemblies available. While we selected the genome with the greatest mapping efficiency to our laboratory strain, a chromosome-scale genome with better gene annotation would probably improve the mapping and help to recover more reads. *Aurelia* is not generally considered a model organism largely due to the inability to close the animal's life cycle in a laboratory setting. Access to early developmental stages is very limited and gametes or larvae can only be harvested if animals are captured from their natural environment, and consequently gene manipulation methodologies are limited. Further, cell dissociations can only be achieved through enzyme-based approaches due to large amounts of mesoglea that is difficult to remove from cell dissociations. Immune response to the presence of degraded extracellular matrix components could elicit a transcriptomic stress response, broadly influencing gene expression values across the dataset. The nature of the tissues also renders in situ hybridizations challenging, particularly with lowly expressed genes.

Material and Methods

Aurelia coerulea (formerly sp.1) culture:

The lab strain was obtained from the collection of Dr. Harms (University of Hamburg approx. 25 years ago). Genomes of several *Aurelia* strains have recently been published and so sequenced reads were first mapped to all available *Aurelia* genomes, with gene annotations extended by 1000bp towards the 3' end, or until reaching the next gene model in the same orientation (17,68,69). We were able to align 90% of the reads from our libraries to the genome of the San Diego aquarium strain (17), compared to only 40% aligning to the atlantic strain or 84% to the pacific/Roscoff strain (68). CO1 gene trees cluster the Roscoff and SanDiego aquarium strain together with *Aurelia sp.1*, which is now identified as *Aurelia coerulea* (70). From these data we identify our laboratory strain as a member of the *coerulea* species. *Aurelia* scyphistomae were grown on small petri dishes placed in 80ml glass containers filled with ASW (artificial sea water) with the salinity adjusted to 35‰. Polyp cultures were fed with *Artemia* nauplii once a week. Containers were cleaned from algae approximately once a month. Strobilation of polyps was induced through either a shift in temperature of about 2-5°C or an 24h treatment with 10µM Indomethacin (18). Freshly liberated ephyrae were transferred to a 5L beaker. Water motion was induced via plastic propellers attached to a small 12V engine. Ephyrae were fed once a day with freshly hatched artemia nauplii. Small jellyfish were

transferred to a 30L Kreisel Tank and daily fed with *Artemia* nauplii, which were incubated in Selco S.presso (INVE) and diluted isochrysis extract overnight to enrich live food with nutrients. Kreisel Tanks were cleaned once a week.

Single-cell dissociations:

Magnesium-calcium-free artificial seawater (MCF-ASW) was prepared according to (71). Scyphistomae, strobilae, ephyrae, or juvenile medusae were collected separately into 1.5ml tubes, washed 2x with FSASW (filter sterile artificial sea water) and incubated in MCF-ASW, followed by incubation in 300µl Collagenase/Dispase Blend I (Merck) dissolved in 35‰ FSASW at RT. Mechanical force via pipetting was used to disrupt the rest of the tissues and obtain single-cell suspensions after 30 minutes of enzymatic digestion. The suspension was monitored for cell clumps using a compound microscope. To stop the digestion once cell clumps were fully dissociated, the suspensions were put on ice and 0.1% BSA was added (bovine serum albumin). Cell suspensions were transferred to a pre-coated (0.2% PTW followed by extensive washing) 1.5ml tube and pelleted for 5 minutes and 0.4rcf using a pre-cooled centrifuge (10°C). The supernatant was removed, and cells were resuspended in 100µl of BSA (0.1% final concentration) and 100ml MCF-ASW. After resuspension samples were filtered through a Flowmi-filter (40µm). Cell recovery and viability were measured with a Nexelcom cell counter upon incubation with 10µM Fluorescein and Propidium iodide. Finally, quality single-cell suspensions were diluted to 1000 cells / uL and loaded according to kit protocols into microfluidic chips with the target cell capture of 7000 cells (10X Genomics, three-prime RNA, vs. 3). Tagged single cell RNA was then processed into a sequencing library following the manufacturer protocol. The resultant single-cell libraries were sequenced with the Illumina platform according to the 10X genomics specifications for paired end sequencing.

Transcriptome mapping:

The best matching genome mapping had only 40% of reads assigned to an interpretable gene model. To improve data recovery the following approach was used: Bulk RNAseq libraries generated in house were assembled with the Trinity assembler v2.8.4 using standard parameters. Additionally, the same libraries were aligned to the San Diego reference genome (17) using STAR v2.3.7a with standard parameters. Single cell RNAseq libraries were also aligned to the same reference genome using cellranger v6.1.2. Final bam files were merged, and gene models were called with stringtie v2.2.1 using standard parameters. This way, three transcriptomes were available: reference guided transcriptome (stringtie), *ab initio* transcriptome (San Diego transcriptome) and a denovo transcriptome produced by trinity. All three transcriptomes were merged, and redundant transcripts were removed from the

assembly using two rounds of clustering with cd-hit v4.8.1. In the first round, all transcripts that were completely contained in another larger transcript were removed. In the second round, CDS of representative transcripts were predicted with transdecoder v5.5.0 prior to clustering. Redundant CDS were removed following the following criteria: 1) at least 90% of the CDS was completely contained in another CDS; 2) minimum sequence identity of 99%. Finally, the quality of the transcriptome assembly was assessed by BUSCO completeness percentage, BUSCO duplication percentage and RNAseq mapping efficiency. Mapping to this final transcriptome increased confidently mapped genes to 60% (**Data S2.1**)

Functional annotation

Final transcripts were functionally characterised using different databases including Pfam, Uniprot, NCBI NR, InterproSites, TMHMM, SignalP and eggnogmapper. Putative names were assigned based on sequence homology to Uniprot best blast hit or NCBI NR best blast hit (**Data S2.2**).

Single-cell analysis

Raw sequence reads were first processed through the cell ranger pipeline with forced recovery of 7000 cells and mapped using a custom mapping tool. Genes are named here according to their best blast hit whenever possible. UMI-collapsed count matrices were imported into R for further processing with the Seurat package (vs. 4.3) (19,72,73) [GenerateAaAllData.R https://github.com/technau/AureliaAtlas]. Expression matrices from individual samples were first filtered to include only samples with at least 750 unique reads. Samples were further filtered for minimal genes (sample range: 350-500), and putative doublets were removed by eliminating outliers with high count values (sample range: 4000-40000). The data was then log normalised with a scale factor of 5000 reads (Seurat::NormalizeData), and top 2000 variable genes were identified for each sample independently (Seurat::FindMarkers). To avoid clustering according to cell cycle state, the resultant variable genes were filtered to exclude those involved in regulation of the cell cycle. Scaled gene expression of this combined set was used as input to the RunPCA function prior to data integration. Samples were then integrated with a reciprocal PCR approach, using the Seurat::FindIntegrationAnchors and Seurat::IntegrateData functions.

The resultant integrated dataset was then used for analyses of cell-state composition following a standard Seurat workflow. Briefly, the integrated data was scaled and used to identify the first 22 principal components with standard deviation of >2. These 22 components were used to generate a nearest neighbour graph (Seurat::FindNeighbors with annoy.metric = ‘cosine’,

k.param = 10) from which cells were clustered with a resolution of 0.1. To visualise the clusters, all cells were projected in a two-dimensional space (UMAP). Each population was then re-analyzed separately, and distinct cell states and/or cell types from each population were annotated. Cell cluster identity was assigned semi-automatically as described in (13). Gene sets used to annotate the single libraries are found in Supplemental **Data S2**. These annotations were then imported into the full dataset for further exploration. To explore data from the perspective of the life history stage, samples of a similar stage were grouped together (polyp = 3, strobila = 2, ephyra = 2, medusa = 5). Cell state tree was calculated by first extracting the embeddings corresponding to the first 50 components computed with Seurat::RunPCA, generating a distance matrix with the euclidean method (stats::dist), and subsequently generating a neighbour joining tree with the ape R package (vs5.8: ape::nj). Upregulated marker genes for each cluster (or stage of the life cycle) were calculated with the Seurat 'FindAllMarkers' function, returning markers that show a fold-change equal or greater than 0.6, detected in at least 30% of the cluster, and having a p-value threshold of >=0.0001. For presentation purposes, smoothed gene expression plots were generated by first imputing missing data, while retaining biological zeros with the SeuratWrappers::RunALRA function ((74)). Scripts for running all analyses and generating the figures found in the paper can be found on our GitHub page: Linketal.Ac.generate.data_publish.R; Linketal.Ac.analyse.alldata_publish.R; Linketal.Ac.analyse.subsets_publish.R; Linketal.Ac.generate.figures_publish.R : https://github.com/technau/AureliaAtlas.

Comparison with *Nematostella*

For direct comparisons with *Nematostella*, we first ran OMA (24) with default parameters on the two transcriptome references. The data matrices for the *Nematostella* single cell atlas of cnidocytes and neuroglandular cells is available at https://sea-anemone-atlas.cells.ucsc.edu under the “Nv2 Atlas/Cnidocyte subset” and “~/All Neuroglands subset” subdirectories. To generate neural expression profiles for *atoh8* orthologs we first removed all “S”, “GD”, and “.early” cell clusters, leaving only the set of putative neural profiles. To compare cnidocyte specification pathways, we removed the following cell clusters from the *Nematostella* dataset: “early”, “gast.1”, “planula.spir”, “planula.nem” and “planula.mat”. For the *Aurelia* cnidocytes we removed the clusters “early”, “medusa.1”, and “medusa.2”. The subset of each data matrix for the set of putative 1:1 orthologs was extracted from both species and aligned using the harmony package (vs.1.2) (25). The cell state tree including both species was then calculated using 20 harmonized principal components and the neighbour-joining algorithm as described above. Support for the recovered nodes was evaluated with the ape::boot.phylo function with 1000 replicates.

Immunohistochemisty and phalloidin:

Animals from different life-stages were relaxed at 4°C followed by a fixation step via 4%PFA at RT for 2h and washed two times in 0.2% PTW after fixation with a minimum washing time of 1/2h. Further, they were blocked using 3% (w/v) BSA (bovine serum albumin) for 2 minutes and incubated in a 1:50 PHA Alexa568/0.2% PTW and 1:1000 DAPI solution for 2h at RT in the dark. After staining they were washed 3x in 0.2% PTW, incubated in vectashield (mounting medium) at RT for 1/2h, mounted on glass slides and imaged using a confocal microscope (Leica SP8). In case of immunostaining, both the sample and the primary antibodies (βTubulin (Sigma Sc-5274) or FMRFamide (Millipore AB15348)) were blocked (1:300) in 0.1% FCS (fetal calf serum) for 1h at RT (sample) or at 4°C (AB). Afterwards the samples were incubated in the pre-blocked antibodies for 24h at 4°C and washed 2x with 0.2% PTW (1/2h/wash). Secondary antibodies (Goat α rabbit, life technologies; Goat α mouse/rat, Invitrogen; Alexa488) were blocked for 1h in FCS (1:1000) at 4°C and washed. The sample was incubated in the secondary antibodies for 20h (4°C). After incubation treatment followed the same steps as after PHA and DAPI stains.

Cloning of genes and preparation of in situ hybridization probes.

RNA was extracted from single-cells, starting with a digestion of animal tissues form different life-stages in 500ml of Animal Free Collagenase/Dispase Blend I (Merck) dissolved in 35‰ FSASW (filter sterile artificial sea water) prepared with DEPC treated water. Single cells were pelleted with 0.4rcf at 4°C for 5 minutes after digestion. The pellet was resuspended, and RNA extracted from single cells by following a standard Trizol extraction protocol (86). Primer sequences were designed using the primer3 website and marker gene sequences amplified from cDNA obtained from extracted mRNA (**Data S2**). Genes were cloned using a pJET1.2 blunt vector cloning kit according to the manufacturer. Plasmids were transformed into *E. coli* and bacteria grown on ampicillin treated agar plates afterwards. Minipreps were performed using a peqGold miniprep kit (VWR). WMISH probes were generated via a SP6 or T7 transcription kit (NEB) and prospectively labelled with either Fluorescein or DIG-UTP. WMISH or FISH were performed using a standard protocol with Hybe buffers based on Urea adapted to the needs of Aurelia tissues (75). Commercial fluorescein/POD and DIG/POD antibodies and buffers were used according to the manufacturer's protocol (perkin elmer) for the performance of FISH.

Literature Cited

1. Pisco AO, Tojo B, McGeever A. Single-Cell Analysis for Whole-Organism Datasets. Annu Rev Biomed Data Sci. 2021 Jul 20;4(1):207–26.

2. Bump P, Lubeck L. Marine Invertebrates One Cell at A Time: Insights from Single-Cell Analysis. Integr Comp Biol. 2023 May 15;icad034.

3. Ryan JF, Burton PM, Mazza ME, Kwong GK, Mullikin JC, Finnerty JR. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: Evidence from the starlet sea anemone, Nematostella vectensis. Genome Biol. 2006;7(7):R64.

4. Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, et al. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc R Soc B Biol Sci. 2009 Dec 22;276(1677):4261–70.

5. Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, et al. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature. 2011 Feb;470(7333):255–8.

6. Collins AG, Schuchert P, Marques AC, Jankowski T, Medina M, Schierwater B. Medusozoan Phylogeny and Character Evolution Clarified by New Large and Small Subunit rDNA Data and an Assessment of the Utility of Phylogenetic Mixture Models. Collins T, editor. Syst Biol. 2006 Feb 1;55(1):97–115.

7. Helm RR. Evolution and development of scyphozoan jellyfish. Biol Rev. 2018 May;93(2):1228–50.

8. Helm RR, Tiozzo S, Lilley MKS, Lombard F, Dunn CW. Comparative muscle development of scyphozoan jellyfish with simple and complex life cycles. EvoDevo. 2015 Dec;6(1):11.

9. Piraino S, De Vito D, Schmich J, Bouillon J, Boero F. Reverse development in Cnidaria. Can J Zool. 2004 Nov 1;82(11):1748–54.

10. Nakanishi N, Yuan D, Hartenstein V, Jacobs DK. Evolutionary origin of rhopalia: insights from cellular-level analyses of Otx and POU expression patterns in the developing rhopalial nervous system: Otx and POU expression patterns in Aurelia. Evol Dev. 2010 Jul 2;12(4):404–15.

11. Kakinuma, Y. An experimental study of the life cycle and organ differentiation of Aurelia aurita. Bull Mar Biol Stn Asamushi. 1975;15:3–6.

12. Nakanishi N, Hartenstein V, Jacobs DK. Development of the rhopalial nervous system in Aurelia sp.1 (Cnidaria, Scyphozoa). Dev Genes Evol. 2009 Jun;219(6):301–17.

13. Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, et al. Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages. Cell Rep. 2022 Sep;40(12):111370.

14. Cole AG, Steger J, Hagauer J, Denner A, Ferrer Murguia P, Knabl P, et al. Updated single cell reference atlas for the starlet anemone Nematostella vectensis. Front Zool. 2024 Mar 18;21(1):8.

15. Babonis LS, Enjolras C, Ryan JF, Martindale MQ. A novel regulatory gene promotes

novel cell fate by suppressing ancestral fate in the sea anemone *Nematostella vectensis*. Proc Natl Acad Sci. 2022 May 10;119(19):e2113701119.

16. Gemmell BJ, Du Clos KT, Colin SP, Sutherland KR, Costello JH. The most efficient metazoan swimmer creates a 'virtual wall' to enhance performance. Proc R Soc B Biol Sci. 2021 Jan 13;288(1942):20202494.

17. Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, et al. The genome of the jellyfish Aurelia and the evolution of animal complexity. Nat Ecol Evol. 2018 Dec 3;3(1):96–104.

18. Kuniyoshi H, Okumura I, Kuroda R, Tsujita N, Arakawa K, Shoji J, et al. Indomethacin Induction of Metamorphosis from the Asexual Stage to Sexual Stage in the Moon Jellyfish, *Aurelia aurita*. Biosci Biotechnol Biochem. 2012 Jul 23;76(7):1397–400.

19. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 May;36(5):411–20.

20. Richards GS, Rentzsch F. Regulation of *Nematostella* neural progenitors by SoxB, Notch and bHLH genes. Development. 2015 Oct 1;142(19):3332–42.

21. Lin HC, He Z, Ebert S, Schörnig M, Santel M, Nikolova MT, et al. NGN2 induces diverse neuron types from human pluripotency. Stem Cell Rep. 2021 Sep;16(9):2118–27.

22. Sur A, Renfro A, Bergmann PJ, Meyer NP. Investigating cellular and molecular mechanisms of neurogenesis in Capitella teleta sheds light on the ancestor of Annelida. BMC Evol Biol. 2020 Dec;20(1):84.

23. Deryckere A, Styfhals R, Elagoz AM, Maes GE, Seuntjens E. Identification of neural progenitor cells and their progeny reveals long distance migration in the developing octopus brain. eLife. 2021 Aug 24;10:e69161.

24. Altenhoff AM, Train CM, Gilbert KJ, Mediratta I, Mendes de Farias T, Moi D, et al. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. Nucleic Acids Res. 2021 Jan 8;49(D1):D373–9.

25. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289–96.

26. Sierra NC, Gold DA. The evolution of cnidarian stinging cells supports a Precambrian radiation of animal predators. Evol Dev. 2024 Mar;26(2):e12469.

27. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, et al. Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. Science. 2019 Jul 26;365(6451):eaav9314.

28. Kraus JEM, Fredman D, Wang W, Khalturin K, Technau U. Adoption of conserved developmental genes in development and origin of the medusa body plan. EvoDevo. 2015 Dec;6(1):23.

29. Cole AG, Jahnel SM, Kaul S, Steger J, Hagauer J, Denner A, et al. Muscle cell-type diversification is driven by bHLH transcription factor expansion and extensive effector gene duplications. Nat Commun. 2023 Mar 29;14(1):1747.

30. Nakanishi N, Renfer E, Technau U, Rentzsch F. Nervous systems of the sea anemone *Nematostella vectensis* are generated by ectoderm and endoderm and shaped by distinct mechanisms. Development. 2012 Jan 15;139(2):347–57.

31. Horridge A. The Nervous System of the Ephyra Larva of Aurellia aurita. J Cell Sci. 1956 Mar 1;S3-97(37):59–74.

32. Satterlie RA, Eichinger JM. Organization of the Ectodermal Nervous Structures in Jellyfish: Scyphomedusae. Biol Bull. 2014 Feb;226(1):29–40.

33. Chia FS, Amerongen HM, Peteya DJ. Ultrastructure of the neuromuscular system of the polyp ofAurelia aurita L., 1758 (Cnidaria, scyphozoa). J Morphol. 1984 Apr;180(1):69–79.

34. Matsuno, A., Hisamatsu, T. Electron microscopy on striated muscles in the ephyra of Aurelia aurita. Mem Fac Sci Shimane Univ. 1982;(16):61–7.

35. Zimmerman KL, Jamshidi AD, Buckenberger A, Satterlie RA. Organization of the subumbrellar musculature in the ephyra, juvenile, and adult stages of *Aurelia aurita* Medusae. Invertebr Biol [Internet]. 2019 Sep [cited 2023 Aug 7];138(3). Available from: https://onlinelibrary.wiley.com/doi/10.1111/ivb.12260

36. Ottmann D, Leyva L, Reglero P, Prieto L, Alvarez I. Ephyrae and metaephyrae of *Pelagia noctiluca* : stage determination, morphometry and shrinkage. J Plankton Res. 2021 Sep 21;43(5):725–31.

37. Steinmetz PRH, Kraus JEM, Larroux C, Hammel JU, Amon-Hassenzahl A, Houliston E, et al. Independent evolution of striated muscles in cnidarians and bilaterians. Nature. 2012 Jul;487(7406):231–4.

38. Tanaka H, Ishimaru S, Nagatsuka Y, Ohashi K. Smooth muscle-like Ca2+-regulation of actin–myosin interaction in adult jellyfish striated muscle. Sci Rep. 2018 May 17;8(1):7776.

39. Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé MP, Renno J, et al. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. Cell. 2018 May;173(6):1520-1534.e20.

40. Brooks W. The life history of the Hydromedusae: a discussion of the origin of the medusae, and of the significance of metagenesis. Soc Nat Hist. 1886;3(12).

41. Marques AC, Collins AG. Cladistic analysis of Medusozoa and cnidarian evolution. Invertebr Biol. 2005 May 11;123(1):23–42.

42. Salvini-Plawen LV. On the origin and evolution of the lower Metazoa. J Zool Syst Evol Res. 2009 Apr 27;16(1):40–87.

43. Chari T, Weissbourd B, Gehring J, Ferraioli A, Leclère L, Herl M, et al. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia* medusa cell types. Sci Adv. 2021 Nov 26;7(48):eabh1683.

44. Dong Z, Wang F, Liu Y, Li Y, Yu H, Peng S, et al. Genomic and single-cell analyses reveal genetic signatures of swimming pattern and diapause strategy in jellyfish. Nat Commun. 2024 Jul 15;15(1):5936.

45. Fuchs B, Wang W, Graspeuntner S, Li Y, Insua S, Herbst EM, et al. Regulation of Polyp-to-Jellyfish Transition in Aurelia aurita. Curr Biol. 2014 Feb;24(3):263–73.

46. Brekhman V, Malik A, Haas B, Sher N, Lotan T. Transcriptome profiling of the dynamic life cycle of the scypohozoan jellyfish Aurelia aurita. BMC Genomics. 2015 Dec;16(1):74.

47. Li X, Ma X, Chen X, Wang T, Liu Q, Wang Y, et al. The medusa of Aurelia coerulea is similar to its polyp in molecular composition and different from the medusa of Stomolophus meleagris in toxicity. Toxicon. 2022 Apr;210:89–99.

48. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan 6;51(D1):D523–31.

49. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. Nat Rev Neurosci. 2002 Jul;3(7):517–30.

50. Masoudi N, Tavazoie S, Glenwinkel L, Ryu L, Kim K, Hobert O. Unconventional function of an Achaete-Scute homolog as a terminal selector of nociceptive neuron identity. Desplan C, editor. PLOS Biol. 2018 Apr 19;16(4):e2004979.

51. Layden MJ, Boekhout M, Martindale MQ. nematostella vectensis achaete-scute homolog nvasha regulates embryonic ectodermal neurogenesis and represents an ancient component of the Metazoan Neural Specification pathway.

52. Tournière O, Dolan D, Richards GS, Sunagar K, Columbus-Shenkar YY, Moran Y, et al. NvPOU4/Brain3 Functions as a Terminal Selector Gene in the Nervous System of the Cnidarian Nematostella vectensis. Cell Rep. 2020 Mar;30(13):4473-4489.e5.

53. Babonis LS, Enjolras C, Reft AJ, Foster BM, Hugosson F, Ryan JF, et al. Single-cell atavism reveals an ancient mechanism of cell type diversification in a sea anemone. Nat Commun. 2023 Feb 16;14(1):885.

54. Zimmermann B, Montenegro JD, Robb SMC, Fropf WJ, Weilguny L, He S, et al. Topological structures and syntenic conservation in sea anemone genomes. Nat Commun. 2023 Dec 13;14(1):8270.

55. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, et al. The dynamic genome of Hydra. Nature. 2010 Mar;464(7288):592–6.

56. Hartl M, Mitterstiller AM, Valovka T, Breuker K, Hobmayer B, Bister K. Stem cell-specific activation of an ancestral *myc* protooncogene with conserved basic functions in the early metazoan *Hydra*. Proc Natl Acad Sci. 2010 Mar 2;107(9):4051–6.

57. Hartl M, Glasauer S, Valovka T, Breuker K, Hobmayer B, Bister K. *Hydra myc2* , a unique pre-bilaterian member of the *myc* gene family, is activated in cell proliferation and gametogenesis. Biol Open. 2014 May 15;3(5):397–407.

58. Hartl M, Glasauer S, Gufler S, Raffeiner A, Puglisi K, Breuker K, et al. Differential regulation of *myc* homologs by Wnt/β-Catenin signaling in the early metazoan *Hydra*. FEBS J. 2019 Jun;286(12):2295–310.

59. Pascual-Torner M, Carrero D, Pérez-Silva JG, Álvarez-Puente D, Roiz-Valle D, Bretones G, et al. Comparative genomics of mortal and immortal cnidarians unveils novel keys behind rejuvenation. Proc Natl Acad Sci. 2022 Sep 6;119(36):e2118763119.

60. Oakley TH, Plachetzki DC, Rivera AS. Furcation, field-splitting, and the evolutionary origins of novelty in arthropod photoreceptors. Arthropod Struct Dev. 2007 Dec;36(4):386–400.

61. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. Nat Rev Genet. 2016 Dec;17(12):744–57.

62. Kin K, Chen ZH, Forbes G, Schaap P. Evolution of a novel cell type in Dictyostelia required gene duplication of a cudA-like transcription factor. Curr Biol. 2022 Jan;32(2):428-437.e4.

63. Utsumi N, Saiga H. Comparison of the Structure and Expression of otx Genes between Ciona intestinalis and Halocynthia roretzi. In: Sawada H, Yokosawa H, Lambert CC, editors. The Biology of Ascidians [Internet]. Tokyo: Springer Japan; 2001 [cited 2023 Aug 7]. p. 215–8. Available from: http://link.springer.com/10.1007/978-4-431-66982-1_34

64. Sawada H, Yokosawa H, Lambert CC. Sawada H, Yokosawa H, Lambert CC. The Biology of Ascidians. Springer. :215–8; chapter 244.

65. Wang CLA. Caldesmon and Smooth-Muscle Regulation. Cell Biochem Biophys. 2001;35(3):275–88.

66. Webb RC. SMOOTH MUSCLE CONTRACTION AND RELAXATION. Adv Physiol Educ. 2003 Dec;27(4):201–6.

67. Almenar-Queralt A, Gregorio CC, Fowler VM. Tropomodulin assembles early in myofibrillogenesis in chick skeletal muscle: evidence that thin filaments rearrange to form striated myofibrils. J Cell Sci. 1999 Apr 15;112(8):1111–23.

68. Khalturin K, Shinzato C, Khalturina M, Hamada M, Fujie M, Koyanagi R, et al. Medusozoan genomes inform the evolution of the jellyfish body plan. Nat Ecol Evol. 2019 Apr 15;3(5):811–22.

69. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental transition and the evolution of animal body plans. Nature. 2016 Mar 31;531(7596):637–41.

70. Scorrano S, Aglieri G, Boero F, Dawson MN, Piraino S. Unmasking *Aurelia* species in the Mediterranean Sea: an integrative morphometric and molecular approach. Zool J Linn Soc [Internet]. 2016 Oct [cited 2023 Jun 20]; Available from: http://doi.wiley.com/10.1111/zoj.12494

71. Paganos P, Voronov D, Musser JM, Arendt D, Arnone MI. Single-cell RNA sequencing of the Strongylocentrotus purpuratus larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome. eLife. 2021 Nov 25;10:e70416.

72. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun;177(7):1888-1902.e21.

73. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun;184(13):3573-3587.e29.

74. Linderman GC, Zhao J, Kluger Y. Zero-preserving imputation of scRNA-seq data using low-rank approximation [Internet]. Bioinformatics; 2018 Aug [cited 2023 Jun 6]. Available from: http://biorxiv.org/lookup/doi/10.1101/397588

75. Sinigaglia C, Thiel D, Hejnol A, Houliston E, Leclère L. A safer, urea-based in situ hybridization method improves detection of gene expression in diverse animal species. Dev Biol. 2018 Feb;434(1):15–23.

Acknowledgments

The authors acknowledge the LISC (life science computer cluster) for providing infrastructure necessary for the bioinformatics, and CIUS (cell imaging and ultrastructure research) for technical support, equipment, and expertise. We thank the staff of the Tiergarten Schönbrunn (Vienna Zoo: Anton Weissenbacher, Roland Halbauer) and Haus des Meeres (Vienna Aquarium: Daniel Abed-Navandi) for sharing their expertise in cnidarian culture, and the UCSC Cell Browser (Maximilian Haeussler and Brittney Wick) for formatting and hosting these data for interactive exploration.

Funding: This research was funded in whole or in part by the Austrian Science Fund (FWF) P34970 (UT) and P31018 (AGC). BW was funded by NIH/NIMH RF1MH132662, CIRM DISC0-14514 and NIH/NHGRI U24HG002371. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

Author contributions:

Conceptualization: OL, AGC, UT

Methodology: OL, JDM, BZ, AGC

Investigation: OL, SMJ, BZ, KJ, JK, JDM, AGC

Supervision: AGC, UT

Writing—original draft: OL, AGC

Writing—review & editing: AGC, UT

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Raw sequence data will be deposited in the GEO database (https://www.ncbi.nlm.nih.gov/geo/), and the analysed data matrix is available on the UCSC Cell Browser (https://jellyfish-atlas.cells.ucsc.edu/). Scripts for analysing the data and generating the figures in this manuscript can be found on our GitHub page (https://github.com/technau/AureliaAtlas).

Figure Captions:

Fig. 1 Single cell transcriptome atlas covering the polyp-medusa life cycle transition demonstrates shifts in genome usage between key life history stages. (**A**) Twelve single cell transcriptomic libraries were generated from cell dissociations of eight sequential life history stages and four tissue compartments from a sub-adult medusa. Samples are color-coded according to their life-cycle. In the center an UMAP cell plot of the full dataset illustrates the distribution of cells from each library plotted with the polyp samples on top (left) and medusa samples on top (right). (**B**) Distribution of genome use across the life cycle. Number of gene models (>3/10 reads) detected for cells binned according to stage, downsampled to 8000 cells representing each stage, with 100 replicates. (**C**) Venn Diagram illustrates the overlap of detected genes from a single replicate of (B). (**D**) Differentially upregulated genes between life cycle stages, represented as a dotplot. (**E**) Over-represented GO-terms present in the up-regulated gene set highlights prostaglandin synthesis in polyps and overabundance of muscle-related genes in the medusa. (**F**) Distribution of cell type states across all life cycle stages plotted as a 2d UMAP with cells coloured according to their population identity (left) and as a fraction of all cells captured (right). (**G**) Expression of genes not detected in either polyp nor scyphozoa in figure (C), filtered to include only genes up-regulated in at least 5 % of any cell state in the dataset. The number of genes in each cell state is quantified above. Cell partitions are as in (F) and cell states are as in Figure 2. Medusa-specific cell states are highlighted with black outline. * indicates the *aurelin* gene with the highest read count of all medusa-specific genes.

Fig. 2 Integrated cell atlas permits identification of cell states across the life cycle. (**A**) Percent of identified cell type-states across each life history stage and medusa tissue. Epithelial sub-states are not included. Colour code as in (B) and (C) (**B**) Cell state tree illustrating the principal cell-type families present in the dataset. Medusa-specific cell states are highlighted with arrowheads. (**C**) Differentially expressed transcription factors across all cell states in the atlas. All transcription factors are colour-coded according to type, colours as in (D). Medusa-specific cell states are highlighted with black gridlines (**D**) Transcription factor families that are over or under-represented amongst up-regulated genes compared to their distribution in the genome; only families that differ by at least 2% of all expressed genes with DNA-binding domains are illustrated.

Fig. 3. Cnidocyte specification pathways indicate two major capsule types. (**A**) Cnidocyte trajectory (UMAP) and (**B**) distribution (bar chart). Inset: location of subset within the full dataset. (**C**) Neighbour-joining cell state tree of cnidocyte profiles from *Aurelia* and *Nematostella* further indicate putative homologies between early (pink highlight) and late (purple highlight) phases of cnidogenesis. (**D**) Cnidocyte complement in *Nematostella vectensis* adapted from (14). The planula-specific cell clusters are removed. (**E**) Expression profiles of core transcription factor orthologs in *Aurelia* (top) and *Nematostella* (bottom) indicative of both early specification (pink) and late maturation (purple). (**F**) Gene tree of *myc* orthologs from both species. Paralogs show similar expression patterns in both species except for myc3 (green), expressed early in *Nematostella* (top right) and late in *Aurelia* (bottom left). Bootstrap support values in (C) and (F) are indicated at the nodes; black indicates >80%, purple <80>50%, and red nodes are supported by <50% bootstrap.

Fig. 4 Neural cell complement includes two distinct classes, with medusa-specific subtypes in each. (**A**) Distribution and characterization of neural complement, illustrated on an UMAP embedding of the full data (top) and neural subset (bottom) (**B**) Distribution histogram across all samples, medusa-specific cell clusters are highlighted with black outline. (**C**) Two classes can be recognized by differential expression of transcription factors. Class 1 cells express *insulinoma* (*ins1a*) (orange left) and class 2 express *achaete-scute* (*ashA*) and pou4 (*pou4-like1*) (green right). (**D**) Density expression plots illustrating N2-specific α-Tubulin (*tba1-like3*), β-Tubulin (*tbb-like1*), and precursorFMRFamide (*pRFa-like3*) expression. (**E**) An anti-βTubulin antibody stains the motor nerve net of the ephyra. (**F**) Co-staining of anti β-Tubulin and anti FMRFamide visualises cells that are positive for both antibodies, possibly corresponding to cells expressing both tubulin and FMRFamide. The ephyra lappet displays condensed regions of the β-Tubulin positive motor nerve net aligning with the radial canal of the lappet. (**G**) Network of the rhopalia nervous system positive for β-Tubulin. DNN = diffuse nerve net, gan = ganglion, lap = lappet, man = manubrium, MNN = motor nerve net, rho = rhopalia.

Fig.5. Muscle territories in the *Aurelia* polyp. (**A**) Schematic representation of the different smooth muscle territories in the polyp from a lateral view. (**B**) A reference UMAP of the muscle subset with a highlighted portion of smooth (red) muscle cells. (**C-H**) Phalloidin stains actin-rich muscle fields in the animals (arrowheads point to retractor muscles). (**I-K**) Polyp images illustrating polyp morphology (left, arrows point to tentacles, od = oral disc, asterisk = mouth, ft = foot, pd = pedal disc) and the body column (right, labelled with a red square) to show the location for the gene expression in L-M. (**L-M**) In situ hybridization for two marker genes (*mrlc2; myophilin-like2*) confirms the spatial location of the retractor muscles, tentacle muscles, and muscles at the oral disc of the polyp. te = tentacle, od = oral disc, rm = retractor muscle.

Fig. 6. Establishment of striated muscle field during strobilation. (**A**) Consecutive stages of strobilation from polyp (scyphistoma) to strobila and ephyra. (**B**) Foot region of the strobila stained with phalloidin showing polyp retractor muscles (arrowheads). (**C,D,E**) ISH for marker genes that label either smooth or striated muscles in the strobila. (**F**) Schematic for advanced strobila with labelled muscle fields (red) as well as fragmented retractor muscles (blue). (**G-O**) gene expression patterns of *stmyhc1* in the advanced strobila. (**P-S**) gene expression pattern for the same gene in the detached ephyra. st.m = striated muscles, rm = retractor muscles, mo = ,mouth.

Fig.7: Muscle territories in the ephyra. (**A**) Ephyra detachment from the strobila. (**B**) Schematic of the ephyra body plan illustrating tissue anatomy (left) and striated muscle field (right). (**C-I**) Phalloidin labels the different muscle fields in the ephyra that are composed of either smooth, striated or a mixture of both muscle types. (**J-O**) In situ hybridization for *striated type Myosin Heavy Chain5* (*stmyhc5*), as well as genes shared by both muscle types (*stmyhc1, mrlc2*) identified from single-cell RNAseq data. (**P**) An expression plot of striated muscle specific genes (*stmyhc5*) as well as genes that share their expression profile between smooth and striated muscles (*mrlc2, stmyhc1*). man = manubrium, gf = gastric filaments, rc = radial canal, st.m = striated muscle, boxes color code: red = (A) detached ephyra and (B) corresponding scheme of ephyra morphology; yellow (J,K, P); green (M, P); blue (N,O,P); gene expression profile and corresponding expression from single-cell data shown in a dotplot

Fig.8: Shared contractile complex between smooth and striated muscle cells. (**A**) Dotplot of gene expression for muscle-type specific gene sets. (**B**) Schematic shows main molecular

components for myosin-based contraction in bilaterians. (**C**) Hypothesis of muscle contractile apparatus in Scyphozoa.

Supplementary Material

Figures:

Fig. S1A) GO-term enrichment for both biological processes and molecular function represented by the gene set up-regulated in the medusa-derived samples compared to other life cycle stages. **B**) Distribution of individual samples in the dataset illustrated on the UMAP.

Fig. S2: Cell cycle dynamics across the life cycle. (**A-D**) Polyp showing overall enrichment of Edu positive cells in the mouth and tentacle region (n=3). (**E-H**) Advanced strobila showing Edu positive cells within the stacks of developing ephyrae. (**I-P**) Liberated ephyra: Edu positive cells are restricted to the distal parts of the lappets, including the rhopalia (I), gastric cavity and filaments (J,K), as well as within the margin of the manubrium (M,N) and clusters of cnidocytes (O) within the exumbrella (n=3). D,H,L,P depict schematics of the different life-stages from either lateral (D,H) or oral (L) and aboral (P) view. mo:mouth; tb: tentacle base; rho: rhopalia; lap: lappet; gf: gastric filament; man: manubrium; cni.c: cnidocyte clusters. Images were taken with a confocal microscope (Leica SP8) at the imaging core facility (CIUS) of the University of Vienna.

Fig. S3: Digestive gland population. **A**) Bar plot showing distribution of digestive gland cells across life cycle stages and medusa tissue samples. **B**) UMAP distribution of identified clusters. **C**) Gene expression dot plot showing expression profile of top 3 differentially expressed genes for each cluster, separated by life cycle contribution. Full gene list is available in supplementary data S1.7

Fig. S4: Mucin gland population. **A**) Bar plot showing distribution of cell states within the mucin gland cluster across life cycle stages and medusa tissue samples. **B**) UMAP distribution of identified clusters. **C**) Gene expression dot plot showing expression profile of top 3 differentially expressed genes for each cluster, separated by life cycle contribution. Full gene list is available in supplementary data S1.8

Fig. S5: Gene expression dotplot of top 4 differentially expressed genes across all identified cell type states. Full gene list is available in supplementary data S1.4

Fig. S6: Expression of *atoh8* family transcription factors. A) Expression profiles across all identified cell states of each member of the *atoh8* family. **B**) Gene family tree with *Nematostella* (red) and *Aurelia* (blue) paralogs highlighted, illustrating an expansion within *Aurelia*. **C**) Expression profiles of *atoh8* family genes within the neural clusters of both species highlights both broad expression profiles, as well as cell-type specific.

Fig. S7: Gene trees for some key transcription factor families. A) bHLH, **B**) Homeodomain **C**) TBX and **D**) HMG. *Aurelia* gene models are highlighted in blue, and *Nematostella* gene models are highlighted in salmon throughout. Gene clades of interest that are discussed in the main manuscript are highlighted.

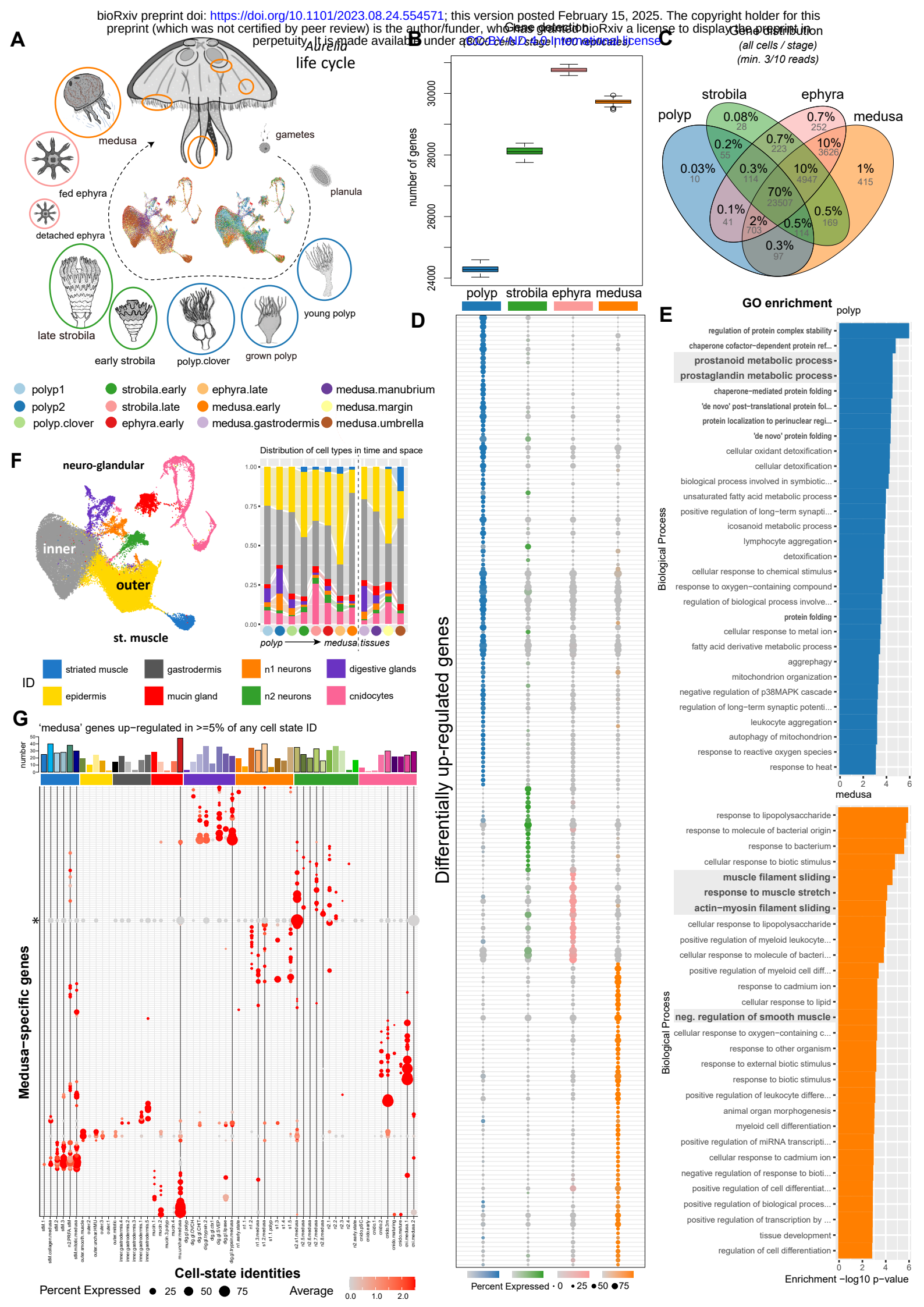
Tables:

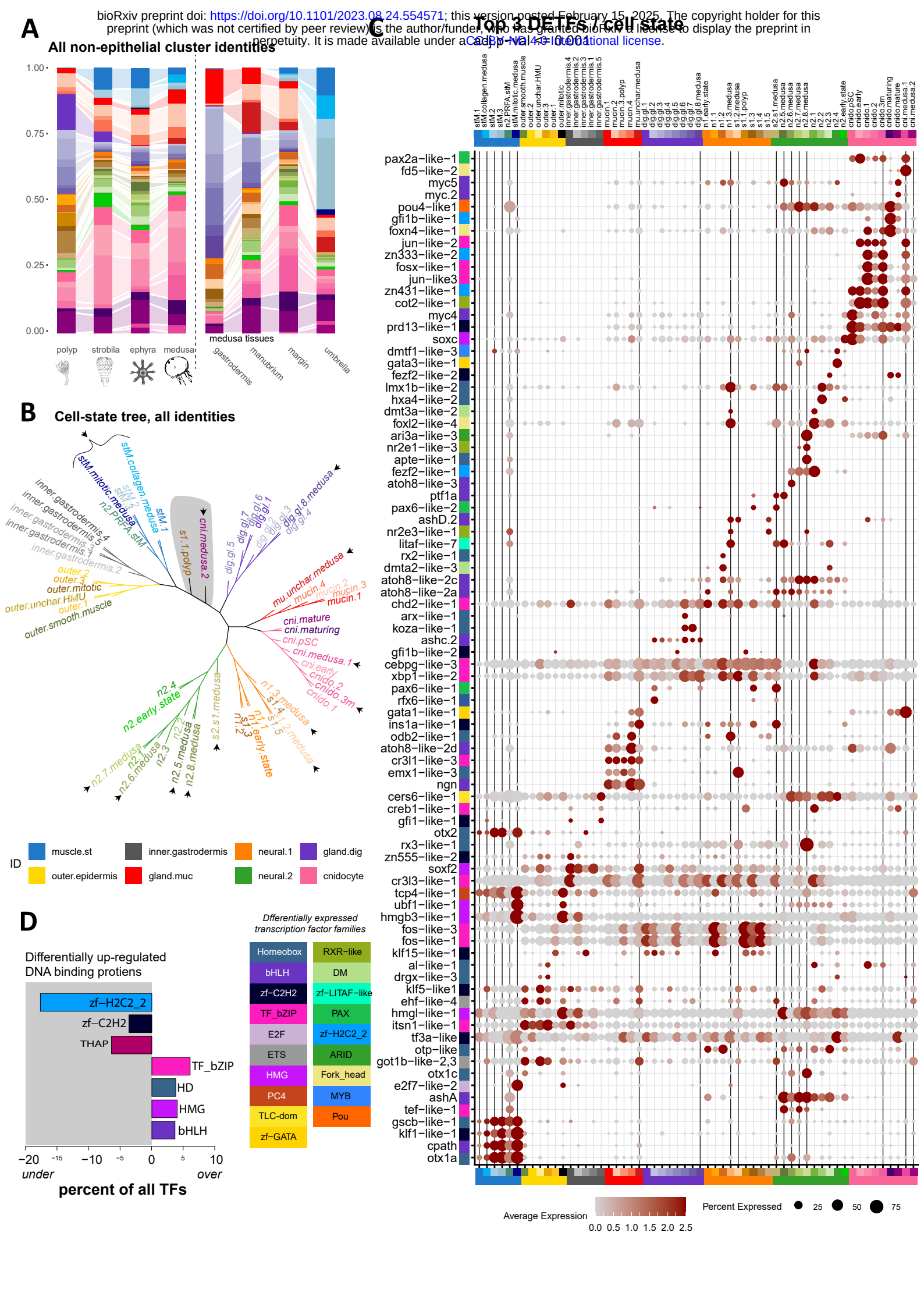
Data S1: Differentially upregulated genes for all cell clusters.

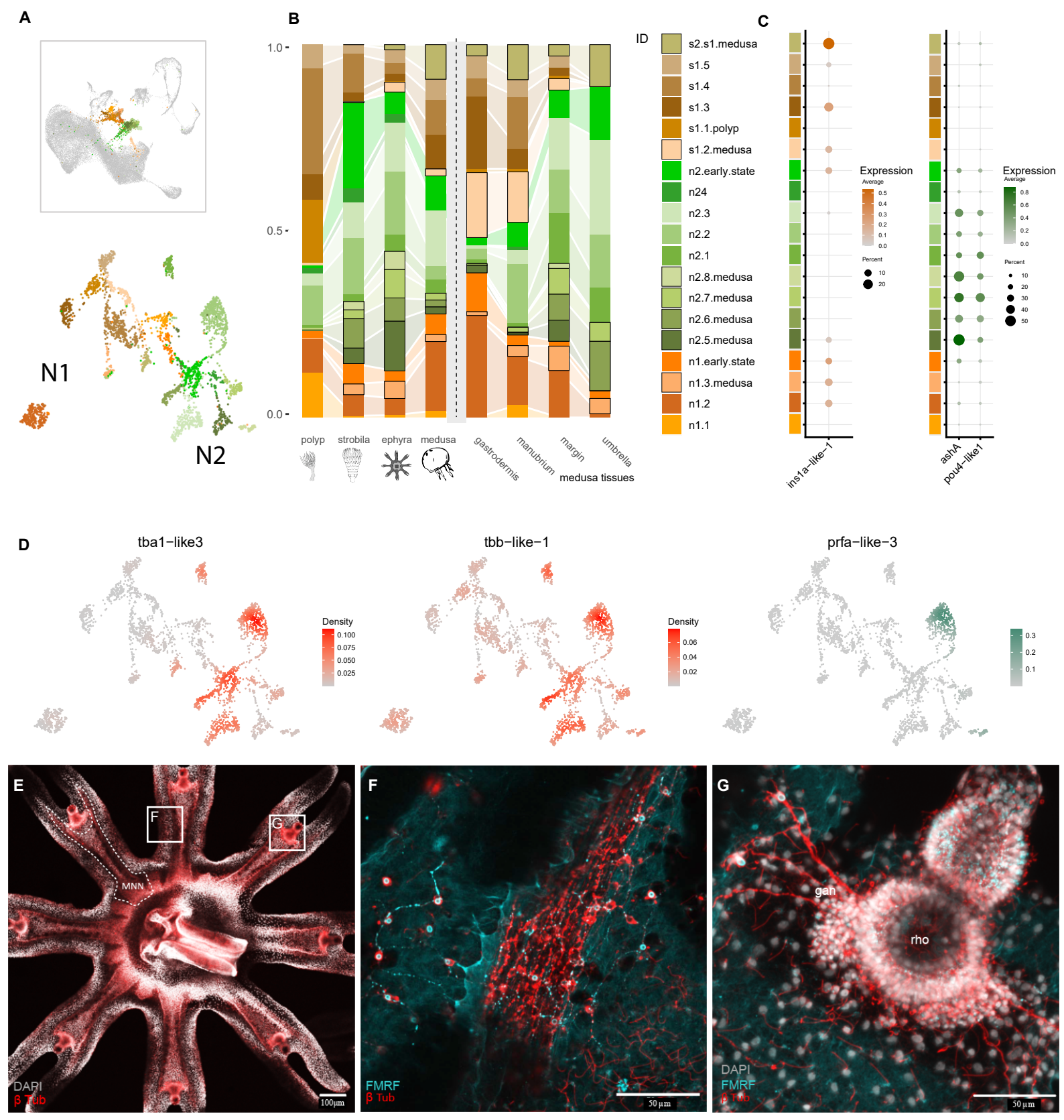
1. Medusa-specific genes from Figure 1C b) filtered list of Medusa-specific genes illustrated in Figure 1G.
2. Life cycle gene set illustrated in Figure 1D. b) DNA binding domain genes (putative transcription factors: pTF).
3. Coarse clustering (Figure 2) b) pTF
4. All.Idents b) pTF illustrated in Figure 2C
5. Epidermis b) pTF
6. Gastrodermis b) pTF
7. Digestive glands b) pTF
8. Mucin gland b) pTF
9. Cnidocytes b) pTF
10. N1 Neurons b) pTF
11. N2 Neurons b) pTF
12. Striated Muscle b) pTF
13. Muscle-specific gene sets

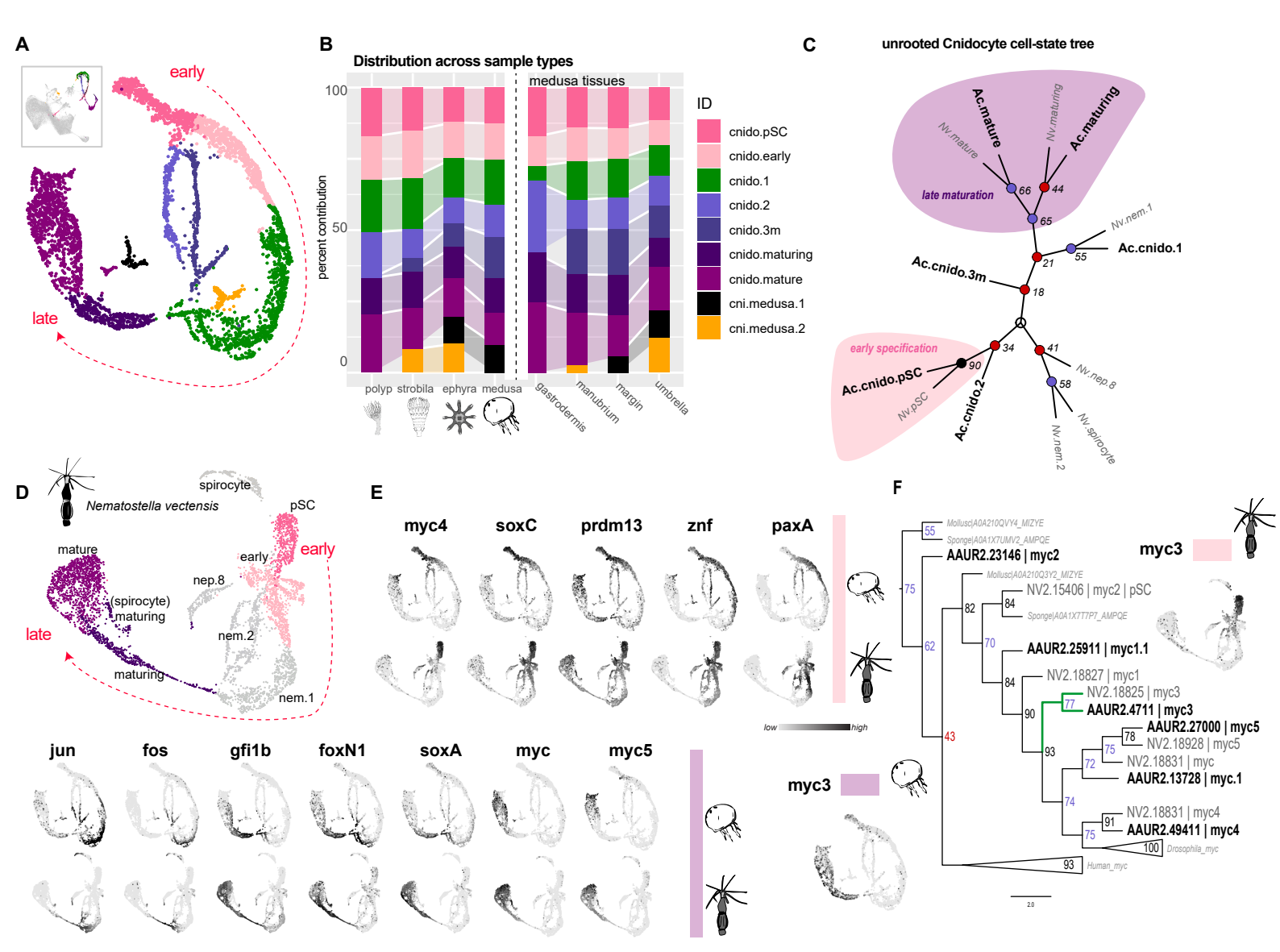
Data S2: Relevant technical information.

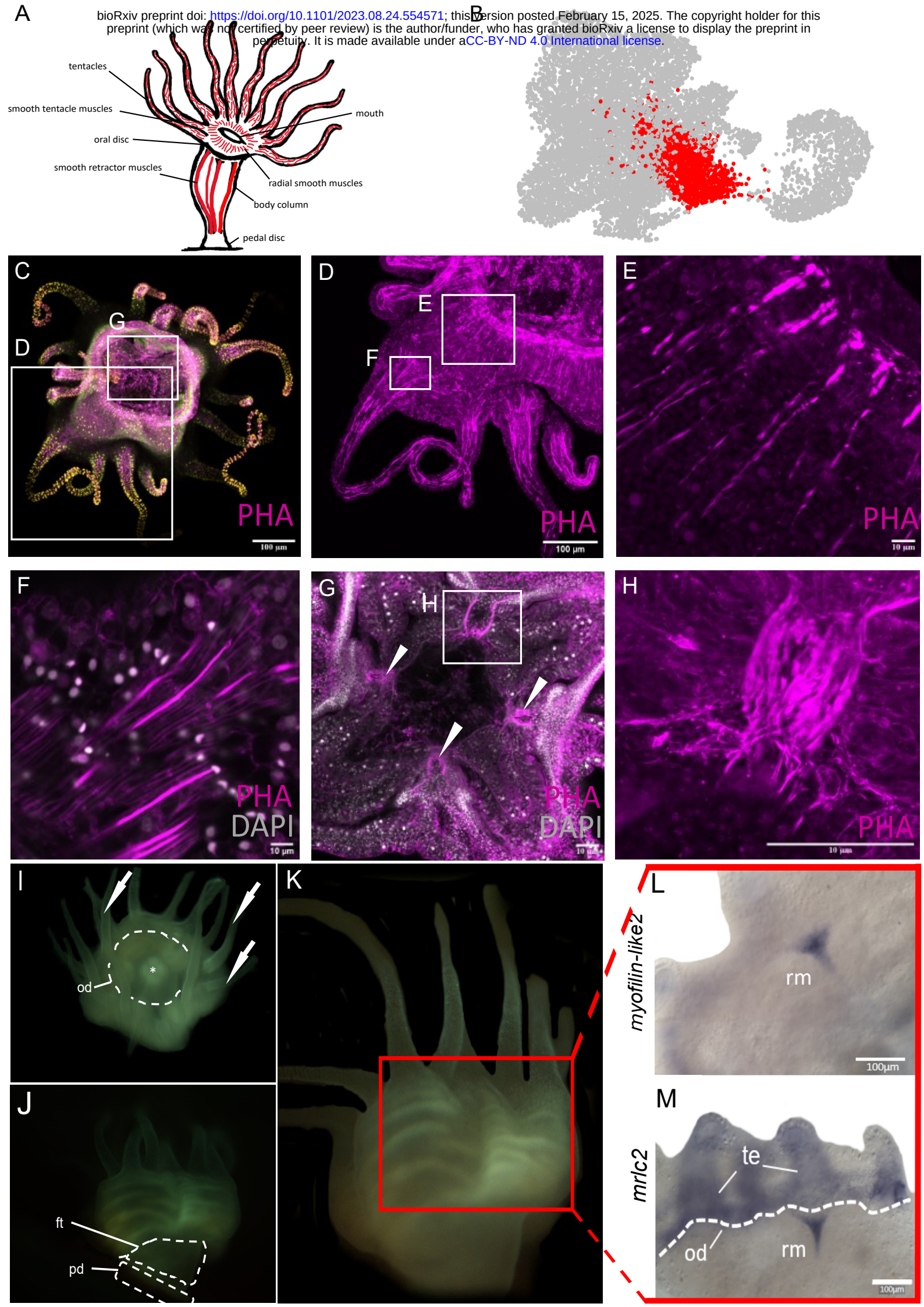
1. Sequence mapping statistics
2. Gene model annotations
3. Coarse clustering
4. Epidermis
5. Gastrodermis
6. Digestive glands
7. Mucin gland
8. Cnidocytes
9. All.Neurons
10. Striated Muscle
11. In situ primer sequences

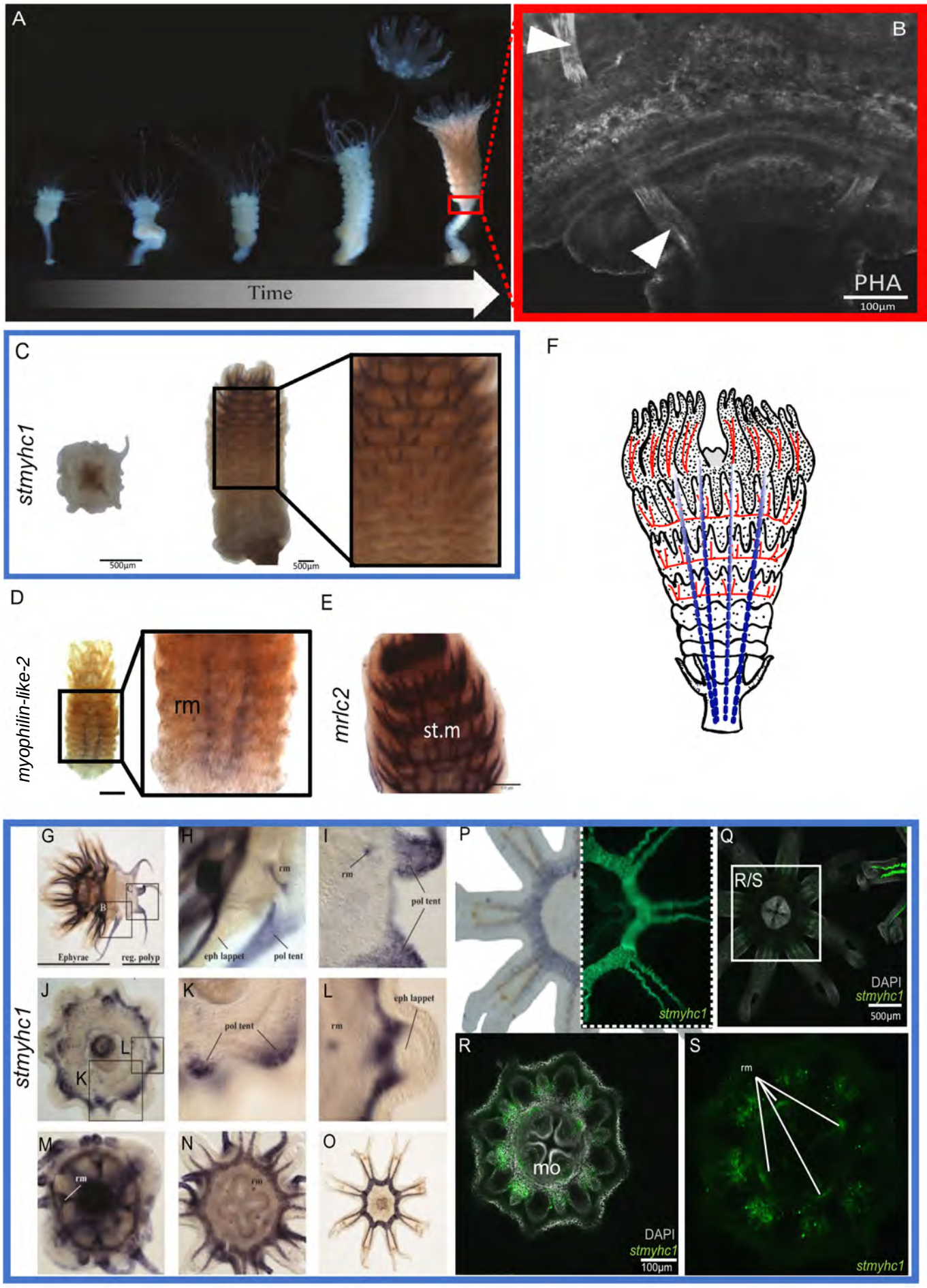


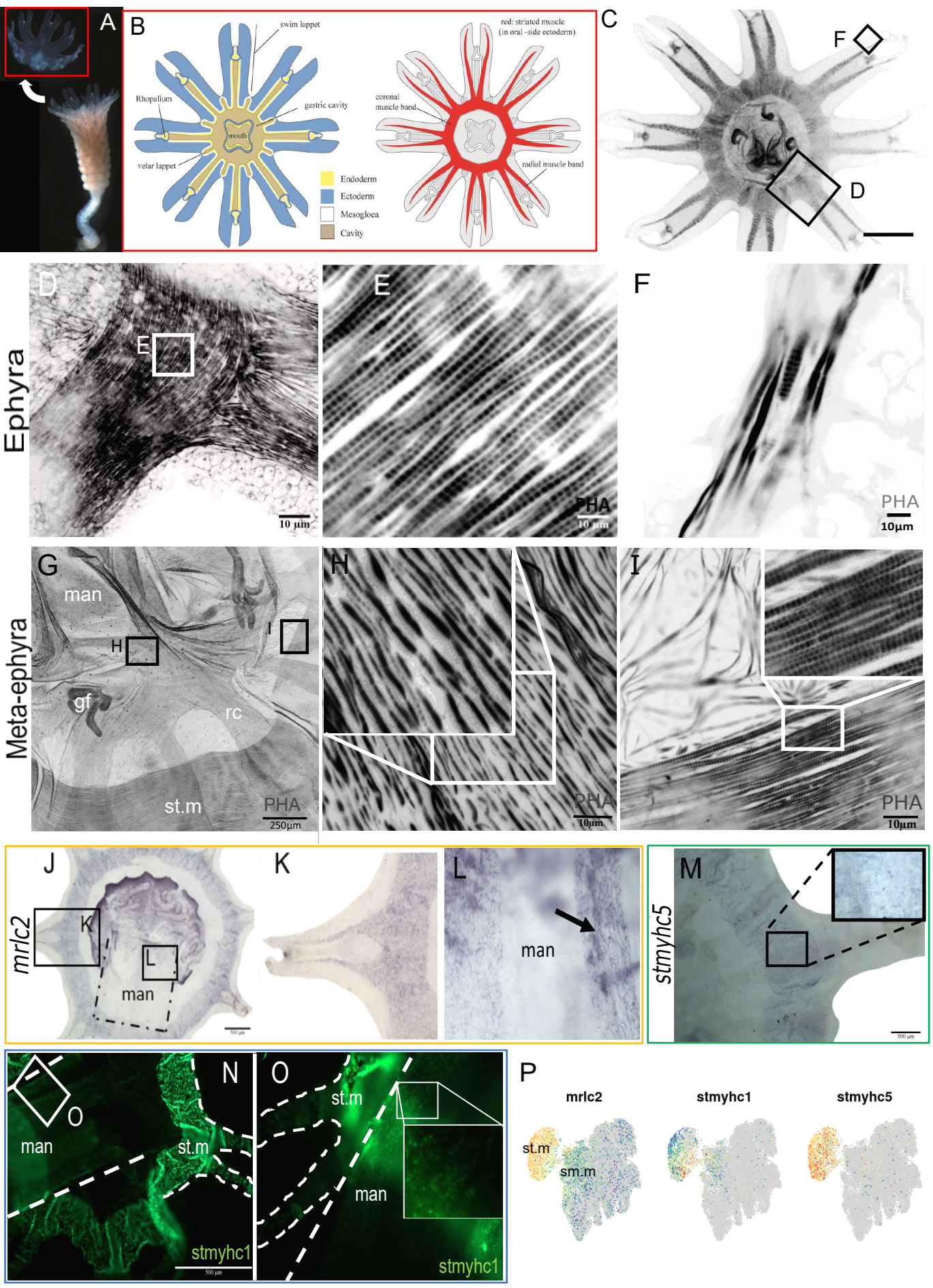



A
N1
N2
B
1.0
0.5
0.0
polyp
strobila
ephyra
medusa
gastrodermis
manubrium
margin
umbrella
medusa tissues
ID
s2.s1.medusa
s1.5
s1.4
s1.3
s1.1.polyp
s1.2.medusa
n2.early.state
n24
n2.3
n2.2
n2.1
n2.8.medusa
n2.7.medusa
n2.6.medusa
n2.5.medusa
n1.early.state
n1.3.medusa
n1.2
n1.1
C
Expression
Average
Percent
ins1a−like−1
ashA
pou4−like1
D
tba1−like3
tbb−like−1
prfa−like−3
Density
E
F
G
MNN
DAPI
β Tub
100µm
FMRF
50 µm
gan
rho


A
early
late
B
Distribution across sample types
medusa tissues
percent contribution
100
50
0
polyp
strobila
ephyra
medusa
gastrodermis
manubrium
margin
umbrella
ID
cnido.pSC
cnido.early
cnido.1
cnido.2
cnido.3m
cnido.maturing
cnido.mature
cni.medusa.1
cni.medusa.2
C
unrooted Cnidocyte cell-state tree
Ac.mature
Nv.maturing
Ac.maturing
Nv.mature
late maturation
66
44
65
Nv.nem.1
Ac.cnido.1
55
21
Ac.cnido.3m
18
early specification
34
41
Nv.nep.8
90
Ac.cnido.pSC
58
Nv.pSC
Ac.cnido.2
Nv.nem.2
Nv.spirocyte
D
Nematostella vectensis
spirocyte
pSC
early
early
nep.8
mature
(spirocyte)
maturing
late
nem.2
maturing
nem.1
E
myc4
soxC
prdm13
znf
paxA
low
high
jun
fos
gfi1b
foxN1
soxA
myc
myc5
myc3
F
myc3
Mollusc|A0A210QVY4_MIZYE
Sponge|A0A1X7UMV2_AMPQE
AAUR2.23146 | myc2
Mollusc|A0A210Q3Y2_MIZYE
NV2.15406 | myc2 | pSC
Sponge|A0A1X7T7P7_AMPQE
AAUR2.25911 | myc1.1
NV2.18827 | myc1
NV2.18825 | myc3
AAUR2.4711 | myc3
AAUR2.27000 | myc5
NV2.18928 | myc5
NV2.18831 | myc
AAUR2.13728 | myc.1
NV2.18831 | myc4
AAUR2.49411 | myc4
Drosophila_myc
Human_myc
55
75
82
84
70
62
84
77
90
43
78
75
93
72
74
91
75
100
93
2.0




A
Time
B
PHA
100µm
C
stmyhc1
500µm
500µm
F
D
myophilin-like-2
rm
E
mrlc2
st.m
G
B
Ephyrae
reg. polyp
H
rm
eph lappet
pol tent
I
rm
pol tent
J
L
K
K
pol tent
L
eph lappet
rm
M
rm
N
rm
O
stmyhc1
P
stmyhc1
Q
R/S
DAPI
stmyhc1
500µm
R
mo
DAPI
stmyhc1
100µm
S
rm
stmyhc1


A
B
swim lappet
Rhopalium
gastric cavity
mouth
velar lappet
Endoderm
Ectoderm
Mesogloea
Cavity
red: striated muscle
(in oral -side ectoderm)
coronal
muscle band
radial muscle band
C
F
D
Ephyra
D
E
10 µm
E
PHA
10 µm
F
PHA
10µm
Meta-ephyra
G
man
H
I
gf
rc
st.m
PHA
250µm
H
PHA
10µm
I
PHA
10µm
J
mrlc2
K
L
man
K
L
man
M
stmyhc5
N
O
man
st.m
stmyhc1
O
st.m
man
stmyhc1
P
mrlc2
stmyhc1
stmyhc5
st.m
sm.m