

# Updated single cell reference atlas for the starlet anemone *Nematostella vectensis*.

**Alison G. Cole**

`alison.cole@univie.ac.at`

University of Vienna: Universitat Wien <https://orcid.org/0000-0002-7515-7489>

**Julia Steger**

University of Vienna: Universitat Wien

**Julia Hagauer**

University of Vienna: Universitat Wien

**Andreas Denner**

University of Vienna: Universitat Wien

**Patricio Ferrer Murguia**

University of Vienna: Universitat Wien

**Paul Knabl**

University of Vienna: Universitat Wien

**Sanjay Narayanaswamy**

University of Vienna: Universitat Wien

**Brittney Wick**

University of California Santa Cruz

**Juan D Montenegro**

University of Vienna: Universitat Wien

**Ulrich Technau**

University of Vienna: Universitat Wien

---

## Research Article

**Keywords:**

**Posted Date:** January 25th, 2024

**DOI:** <https://doi.org/10.21203/rs.3.rs-3854371/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Version of Record:** A version of this preprint was published at Frontiers in Zoology on March 18th, 2024. See the published version at <https://doi.org/10.1186/s12983-024-00529-z>.

# Abstract

## Background

The recent combination of genomics and single cell transcriptomics has allowed to assess a variety of non-conventional model organisms in much more depth. Single cell transcriptomes can uncover hidden cellular complexity and cell lineage relationships within organisms. The recent developmental cell atlases of the sea anemone *Nematostella vectensis*, a representative of the basally branching Cnidaria, has provided new insights into the development of all cell types (1,2). However, the mapping of the single cell reads still suffers from relatively poor gene annotations and a draft genome consisting of many scaffolds.

## Results

Here we present a new wildtype resource of the developmental single cell atlas, by re-mapping of sequence data first published in Steger, Denner, Cole, *et al* 2022 (1) and Cole, Jahnel *et al*, 2023 (3), to the new chromosome-level genome assembly and corresponding gene models in (4). We expand the pre-existing dataset through the incorporation of additional sequence data derived from the capture and sequencing of cell suspensions from four additional samples: 24hr gastrula, 2d planula, an inter-parietal region of the bodywall from a young unsexed animal, and another adult mesentery from a mature male animal.

## Conclusion

Our analyses of the full cell-state complement provide transcriptomic signatures for 127 distinct cell states, of which 47 correspond to neuroglandular subtypes. We also identify two distinct putatively immune-related transcriptomic profiles that segregate between the inner and outer cell layers. Furthermore, the new gene annotation Nv2 has markedly improved the mapping on the single cell transcriptome data and will therefore be of great value for the community and anyone using the dataset.

## Introduction

The establishment of single cell transcriptomics in a wide range of organisms that are usually not accessible to genetic manipulation has opened the avenue to make large-scale comparisons of cell type complexity and evolutionary origin of specific cell types. However, a key aspect in such comparative approaches is the quality and depth of the underlying datasets. This largely depends on the quality of the genome and the corresponding gene annotations. For the sea anemone *Nematostella vectensis*, which is now one of the major model organisms among cnidarians, single cell transcriptomes have been published, which cover the developmental stages from the gastrula to the adult stage (Sebe-pedros, 2018, Steger et al 2022; Cole et al). However, the underlying single cell reads have been mapped on a draft

genome, consisting of numerous scaffolds and a reference transcriptome, called Nve, which does not always include the 3' UTR required for mapping of reads derived from the 10X Genomics platform. As a result, while we artificially extended the gene models in the 3prime direction, some genes may have been erroneously annotated. Our group recently generated a new chromosome-scale assembly and – using PacBio Isoseq reads – an improved, new gene annotation, called Nv2 (4). The new gene annotation requires a new mapping tool to fully explore the single cell dataset.

Here we present a new wildtype single cell transcriptome resource for the anthozoan model *Nematostella vectensis*, based on the re-mapping of sequence data first published in Steger, Denner, Cole, *et al* 2022 (1) and Cole, Jahnel *et al*, 2023 (3), to the new chromosome-level genome assembly and corresponding gene models in (4). We further expand the pre-existing dataset through the incorporation of additional sequence data derived from the capture and sequencing of cell suspensions from three additional samples: 24hr gastrula, 2d planula, and another adult mesentery from a mature male animal. In general, we recapitulate the previously reported features of the dataset, and expand our cluster annotations from the previous publications. We provide an in-depth analysis of the three main tissue layers: the inner gastrodermis, the outer epithelium, and the cells from the boundary between these that contribute to the pharynx and septal filaments. We make the revised analyzed dataset available for public exploration on the UCSC single cell web browser ([sea-anemone-atlas.cells.ucsc.edu](http://sea-anemone-atlas.cells.ucsc.edu)).

## Results and Discussion

We analyzed the re-mapped data in two phases: 1) we merged all samples together to generate the complete dataset; 2) we analyzed each principal data partition separately to arrive at a complete cell-state atlas. We then imported these cell-state identifiers into the merged datasets for further use in data exploration, and we provide here expression data plots separated between the developmental data derived from 18hr post-fertilization to 16day primary polyps (“developmental”), and the adult tissue samples (“adult tissues”). Our results do not differ greatly from the cell type annotations provided in (1). Rather we provide here a finer-grain description of the cell complement present across the life history stages.

We recapitulated our previous analysis and recover the same principal cell partitions with clear maturation signal during embryogenesis within the primary germ layers, and all neuroglandular derivatives present (Fig. 1). Both the inner gastrodermis and the outer epithelial layer show a clear developmental maturation signal evidenced by the separation into two clearly distinct cell partitions within the dataset. This is not likely to be an unresolved batch effect, as we find cells of the mature state contributed from all libraries, while a distinct early profile is associated with cells predominantly from the gastrula stages (Fig. S1, S2, S4, S5).

## Epithelia

The largest partition in our dataset is the ectodermal epithelial cells. The broad underlying groups of ectodermal cells include the embryonic ectoderm, the external body wall epithelium, and the internal pharyngeal ectoderm (Fig. 2A). The ectodermal oral-aboral body axis is established by a gradient of Wnt/ $\beta$ -catenin signaling during early embryonic development, which leads to the specification of three major expression domains: the oral domain, a midbody domain and the aboral domain (5–7). Notably, although the dissociation is thought to abolish all spatial information, the analysis of our data shows that the spatial information along the primary body axis is retained in defined clusters of our gastrula ectoderm subset, with well-known signalling pathways and patterning genes demarcating the oral (*FoxA*), central (*Wnt2*) and aboral (*Six3-6*) domains (Fig. 2B) (8–10). The proportions of these domains are consistent between spatial patterns and our sc-transcriptomic data (Fig. 2B). Furthermore, the embryonic epithelium widely expresses *Ptx1*, a maternally deposited toxin that is important for the protection from predators (**Supplementary Figures and Data 1**; (11,12)). Analysis of the embryonic partition in isolation reveals a cluster of mitotically active cells, cells with aboral and oral profiles, as well as one cluster with a strong ribosomal profile that is present across all libraries and may be a technical artifact (**Fig. S1**). Within the mature epithelial ectodermal partition, we can further differentiate 7 cell states, 4 of which show distinct transcriptomic properties indicative of ciliated cells (*foxJ1*: yellow), progenitor cells (*myc2*: dark grey), actin-enriched (green: ect.3) and a cycling mitotic population (histone-rich) (**Fig. S2, Supplementary Figures and Data 1**).

The aboral pole of the planula larva is characterized by a small region with a long ciliary tuft commonly referred to as the apical organ (ao). It is a transient structure that disintegrates during metamorphosis, and likely preforms sensory functions (13–15). These cells first appear in the 24hr gastrula, are sparsely detected in the 5d sample, and are absent again from the 8d primary polyp samples (Fig. 2E). Early apical tuft cells from the gastrula samples are enriched for *fgf1a*, *FoxN1*, and *vent1-like* expression (*ect.AO.early*: Fig. 2C, **Supplementary Figures and Data 1**). We further identify two apical organ-related cell types corresponding to centrally located apical organ cells (*isx1-like-1* expression: spot) and the surrounding sensory neurons (*Slc* expression: ring) (Fig. 2C, **Supplementary Figures and Data 1**, (15)).

The ectoderm at the boundary with the gastrodermis has unique properties in terms of both its developmental formation, and its contribution to the adult. Cells from this region form the ectodermal portion of the pharynx as well as lining the distal-most portion of the mesentery folds. This territory is enriched in digestive gland cells and has been proposed to be homologous to the bilaterian endoderm (16). In support of this hypothesis, this territory can be identified throughout development by the expression of the transcription factor *foxA*; the vertebrate *foxa2* ortholog is a marker of the definitive endoderm (17,18). Within the *foxA* positive pharyngeal ectoderm and septal filament territory we also further identify 7 cell states, including an additional population of mitotically active cells (histone-rich), a population of early blastopore lip cells, a DLGP5-positive cell state, and a LRWamide-sensitive (LRWamide-receptor expressing) DMBX1-positive cell state (**Fig. S3**).

## Gastrodermis

One of the first cell fate decisions in the embryo is the specification of the germ layers, wherein derivatives of the inner cell layer, commonly called (mes)endoderm (but see (16) for an alternative view), are identifiable in their mature state by high levels of specific soma ferritins (FRIS-like) and the production of collagen filaments, presumably in preparation of the forming mesoglea (**Supplementary Figures and Data 1**). As noted for the ectodermal partition, there is a strong maturation signal evident also in the inner cell layer (Fig. 3A). The early gastrodermis has a unique transcriptomic signature rich in histones genes indicative of active proliferation, and the three identified cell states do not have unique expression profiles (**Fig. S4**). There is an intermediate maturation phase at 2d (green clusters, Fig. 3), wherein also the pharyngeal gastrodermis is detectable (dark red, Fig. 3B). This state is then replaced by differentiation of various cell types that are maintained into the adult. The mature gastrodermis itself is characterized by the lack of an additional differentially expressed gene set when compared to the remaining mesendodermal cell (Fig. 3C). The transcriptomic profile corresponding to the mature gastrodermis of the pharynx appears at 2d, the mature gastrodermis profile appears at day 3, while at 4d the non-muscle mesendoderm is first evident, characterized by the expression of the large glycoprotein *cpg2-like-1*, the lectin *cs/3-like-3*, as well as phospholipase PA2 orthologs (PA2-like9). One cell state is restricted to the mature mesentery samples, and thus could represent the presumptive somatic gonad ("**mesentery**") (Fig. 3B). As the inner layer matures, differentiated cell types emerge. We detected ciliated cells associated with the body wall samples, secretory cells with a gland-like profile (expressing *mucin*), a putatively neural population (expressing *LRWamide-1*), and a distinct immune state which expresses IRF orthologs (expressing *irf1-2b* and *irf1-2c*) (Fig. 3B:D, **Supplementary Figures and Data 1**). These latter cells also express *snailB*, which transitions to a single-cell expression territory post-gastrulation. This cell state is discussed further below. The mature inner cell layer is rich in apolipoporphins, indicative of nutrient storage and mobilization expected of the gastrodermis. Within this layer we can identify one cell state that expresses *twist* and *VAX-EMX-like*, which represents the pharyngeal gastrodermis where these gene have been reported to be expressed (3,16). Like that described for the adult tissue-only samples (3), cells corresponding to the circular musculature, the inter-muscular membrane, and the parietal and early mesentery retractors are identifiable within the dataset (Fig. 3A). At the resolution explored here for clustering, the parietal and early mesentery cells cluster together. Mature mesentery retractor cells are found rather within the retractor muscle cluster itself (Fig. 4).

## Retractor muscle

The retractor muscle profile can be separated into four distinct profiles, two of which are indicative of earlier (TR.1: *mf43*) and later states (TR.2: *gbra5*; *grbg3*) of the tentacle retractor muscle (*nem64*), one of the mesentery retractor (MR.1: *nem24*), and one stable mature cell state that is a convergent profile of both fast muscles types as has been previously described (3) (Fig. 4). While the mature profile is shared across all clusters, the ectodermally-derived tentacle muscle expresses GABA/Glutamate receptors that are absent from the mesentery counterpart. We previously demonstrated that the ectodermal tentacle muscle likely derives from the neuroglandular progenitor (NPC) lineage, as the tentacle retractor cells express *soxB(2)* (also called *SoxB2a*) (3) and populations of *nem64* expressing cells can be found within

the putative stem cell partition (Fig. 6; (1)). The early tentacle profile associated with the differentiated retractor muscle here expresses *wnt1* and is restricted to samples obtained from the pharynx-region of the adult. Interestingly these cells also express *mf43*, a zinc finger known to negatively regulate wnt signaling in other systems (19,20). Previously both FGF (21) and Notch (22) signaling have been demonstrated to play a role in tentacle outgrowth. Wnt signalling is crucial for establishing the oral pole in the embryo and indeed acts as an organizer capable of inducing ectopic head formation, including tentacle outgrowth, in both embryos (5) and regenerating adults (23,24). Wnt expression within tentacle-specific cell types, as shown here for the tentacle muscle cells, indicates a role for wnt signaling in regulating tentacle formation in a cell-type specific manner.

## Cnidocytes

**Figure 5: Cnidocyte specification pathways are recovered.** **A** Cell plot indicating the trajectory clusters of the cnidocyte partition (inset). **B** Distribution of sample contribution to each identified cluster. **C** Specific toxin profiles associated with cnidocyte subtypes. **D** Dotplot of marker expression (box) and differentially up-regulated genes of each cluster. See Supplementary material for full gene lists.

Cnidocyte trajectories described in Steger *et al* 2022 (1) are recovered (Fig. 5). Here, a larger portion of the putative stem cells are included in this cluster (blue cluster: *sox3* positive; Fig. S7, Supplementary Figures and Data 1), reflective of the fuzzy cluster boundaries inherent in these methods. The different cnidocyte sub-types express different members of the membrane attack complex (*mac1* and *mac2*) (Fig. 5C). Updated gene lists representing the expression profiles associated with the distinct phases of the specification trajectories are available here with the updated genome mapping (Supplementary Figures and Data 1).

## pSC and PGCs

We identify two partitions that contain the putative stem cells (pSC) and the primary germ cells (PGCs; Fig. 6A). The latter were previously identified as part of the pSC cluster, and it was suggested that this was the first fate choice facing the stem cell population upon exit from the mitotic cycle (1). Here we include an additional mature mesentery from a male specimen, thereby enriching for spermatogonia. The germ cell cluster is predominantly composed of maturing sperm from this male mesentery, but also contains early oocytes and stem cells (Fig. S9) and derive from the pharynx and female mesentery libraries, consistent with what has been described for early PGC formation in this system (25). This subset contains a full maturation sequence of spermatogenesis and a small cluster of primary oocytes, and early state primary germ cells (PGC), and demonstrates the presence of distinct sets of regulatory factors that may play important roles for gametogenesis in this organism.

Altogether, the stem cell cluster contain ectodermal neuroglandular progenitors (NPC), gastrodermis-derived NPCs (NPC.g), two mitotically active cell clusters indicative of DNA synthesis (mitotic.1) and M phases (mitotic.2) of the mitotic cycle, as well as distinct clusters of early primary germ cells, primary oocytes and maturing sperm (Fig. 6B). NPCs derived from the inner cell layer also express these mitotic markers, further supporting an independent origin of these two cell populations (Fig. 6D). The inner cell

layer NPCs first appear in the 2d planula samples (Fig. 6C), and are identifiable by *six1/2* expression, previously documented to be enriched in the inner cell layer (26).

## Neuroglandular complement

The neural and gland cell types cluster together into a large neuroglandular population, reflective of their close developmental relationships as previously documented. As per (1) here we identify 47 distinct cell states (Fig. 7) which can be broadly separated into *ashC* positive digestive gland cell types (GD: 5 states), *nem64* positive early tentacle retractor cell states (N2.TR: 2 states; (3)), IRF positive putative immune cells (1 state), putative sensory (N1S: 5 states), uncharacterized secretory (S1: 4 states, S2: 4 states, and S3: 1 state), neurons (N: 18 states), and some early progenitor states (7) with elevated *soxB2a* and *soxC* expression (Fig. 7D). As described in (1) neurons can be divided into insulinoma (INSM) positive (N1: 16, including 3 larval states (N1.L)) and INSM negative (N2: 6) populations. These can be further sorted according to germ layer origin, where gastrodermis-derived neurons express *six1-2*, *otx*, and/or *nkx2.2D*. There is only one INSM (-) gastrodermally derived population (N2.g1). All classes of neuroglandular cells reach their peak diversity by the 4d planula. Three larval neural states and one larval S1 state appear at gastrulation and largely disappear once the planula enters the tentacle-bud stage at day 5, although there a few cells in the tentacle and bodywall adult samples that cluster with N1.L3 indicating that this cell type may persist into the adult (Fig. 7B). N1.L1, N1.L2, and S.L1 are instead restricted to the gastrula and planula stages. N1.L2 is likely localized to the oral pole (*six3/6* expression) and may also form part of the apical organ. Gastrodermis-derived neurons appear already in the 2d planula. Only sub-states S2.tll.4 in the tentacles and body wall, and S3 in the mesenteries are restricted to the adult subsets. Early states indicative of tentacle retractor muscle differentiation share expression of *barH* orthologs and *tbx20.3* with inner cell layer neurons N1.g1 (Fig. 7D). While the overall composition of the nervous system in *Nematostella* has been described (27–29), to date there are few isolated neural sub-types that have been fully characterized in terms of their molecular fingerprint. Tourniere *et al* (30) describe the distribution and close relationship between the *pou4*-expressing N2 neurons and the cnidocytes, and provide evidence of subtypes that we identify here that express RFamide (N2.2), glutamate receptor *GRIK2-like-1* (NVE22966; N2.g1: also described here: (31)), and GABA-A receptor (*GBRB3-like-2*; NVE21438), which here is expressed in the presumed early N2 cell state (N2.early, *SoxB(2)* expressing). Further, the same authors also provide characterization of the N1 subset (32) but do not identify specific neural sub-types. Interestingly they also find an insulinoma-positive (N1 class) subset that is immuno-reactive for RFamide, while here we find this peptide restricted to N2 neurons. However, there are other putative RFamide-peptides in the gene set and so there may be other cross-reactive proteins. We also identify here N1S.4 as the *foxQ2d* expressing sensory neuron described in (33). Further characterization of this partition is ongoing and will likely yield interesting hypotheses regarding the early evolution of neural complement in the common ancestor of Cnidaria and Bilateria.

## Mucin gland

The mucin-producing cell type is one of the earlier cell types to arise during development, but little is known about its distribution nor function within the organism. This population expresses multiple mucin-



like polysaccharides and is one of the first secretory cells to arise in development together with *vsx*-positive early neurons, as previously documented by (1). Cell states within this cluster appear to be associated with embryo maturation, with 4 states falling within the developmental series and 3 found within the adult. One cell state appears to be restricted to the mesenteries and has a unique transcriptomic profile, including several specific enzymes (eg. *ats10*, *nas4*, *loxh1*, *asprv*. **Supplementary Figures and Data 1**, Fig. 8). This small sub-type also over-expresses RLRA and a number of heat shock proteins (**Supplementary Figures and Data 1**). Otherwise, the mucin gland cluster is relatively homogeneous and could be considered a single cell type. This cell type is *INSM* positive, reflecting the sister-cell relationship in the hierarchy of specification decisions of the NPCs in the early gastrula (1). The driver of this cell type identity is likely to be the *noto*-related homeobox gene *not2-like-4*, although this has yet to be proven experimentally. While the gel-forming mucin proteins themselves have been demonstrated to pre-date the cnidarian-bilaterian split (34), analyses of the cell types that produce these proteins is still lacking.

## Immune profile

One novel cell state recovered here is that of the putative immune profile. A large partition of cells exhibits this transcriptomic profile, embedded with the epithelial cloud in reduced dimensional space (Fig. 9A). These cells all express elevated levels of *irf1-2a* and *NFKB1* orthologs (Fig. 9E). There are 4 stable states associated with this partition, including one cell state that is enriched for histones and other DNA-modulating factors (**Fig. S12, Supplementary Figures and Data 1**), but otherwise the transcriptomic signature is largely invariant across the partition. We found cells with this signature also within the neuroglandular partition (Fig. 9E **and Supplementary Figures and Data 1**). This transcriptomic state possesses a complex regulatory profile, including a large set of specific transcription factors that includes *nfkb1* (three paralogs), *maf* (three paralogs), and receptors *rlrb* (2 paralogs) and *rlra*. This partition is also enriched in *wnt4* expression, suggesting this *wnt* as an important modulator of this cell state. Altogether the distribution of this transcriptomic state suggests that it could represent a further distinct cell type that arises from the common neural-glandular progenitor population. Alternatively, this cell state could represent an inducible stress response rather than a distinct cell type. Recent studies have described an inducible immune response with a similar molecular composition (i.e.: IRFs, GBPs, OAS, NFKB **Supplementary Data 1**) (35,36). Previous investigations into NFKB function in *Nematostella* have shown that protein expression is localized to a single cell ectodermal pattern embryos (37), and morpholino knock-down experiments suggest that the NFKB pathway plays a role in cnidocyte development (38). While we can detect expression of NFKB orthologs within the cnidocyte lineage, it is significantly more abundant within the ectodermal immune partition (Fig. 9E). However, if this transcriptomic state indeed reflects a previously undescribed immune cell type in sea anemone's that similarly arise from the common neuroglandular stem cell population, perturbations could potentially affect multiple cell derivatives, including the cnidocytes. Brennan *et al* (39) have demonstrated that the NFKB pathway is also active in the nematosomes, ball-like assemblies of stinging cells that are released into the body cavity of *Nematostella*, presumably from the septal filaments. Cells with an amebocyte-like morphology that are capable of phagocytosis have also been isolated from *Nematostella* (40). Of particular interest,

we find a second cell state associated with the gastrodermis which expresses the paralogs *irf1-2b* and *irf1-2c* but otherwise show very little overlap with the ectodermal immune state (Fig. 9E). Similarly in corals two transcriptomic states with molecular signatures indicative of immune function, including *irf1/2* expression, have also been cataloged within the single cell transcriptomic atlas (41). In corals the immune cell states also express multiple secreted proteins, whereas here the more abundant ectodermal immune profile in the anemone does not. The gastrodermal immune state however does express elevated levels of cytoskeletal proteins and peptidases (**Supplementary Data1**). Further experimental investigation of these cell states is warranted to distinguish between these two alternatives: an inducible cell state(s) vs. dedicated immune cell type(s). Nonetheless, it is clear from the cell atlas data presented here that two robust immune-related cell states are present in *Nematostella*, and these are distinct within both germ layers.

## Conclusions

Here we provide an updated wildtype single cell atlas of development and tissue composition for the starlet sea anemone *Nematostella vectensis*. We identify a total of 127 distinct transcriptomic states within the dataset, comprising both unique cell types as well as developmental cell states. This is similar to the first single cell transcriptomic atlas produced for this species, wherein 104 unique transcriptomic states were described from adult tissues and 39 from a larval stage, of which 23 overlapped with the adult complement (2). The neuroglandular partition is transcriptomically by far the most diverse, here accounting for 37% of the cell states, while comprising only 9% of the entire dataset. We can correlate transcriptomic states of the inner cell layer with anatomical structures (different muscle groups, membrane attachment of the mesenteries, inner lining of the pharynx), and similarly the ectodermal layer can be separated into the inner ectodermal or pharyngeal layer and the outer epidermis. However, we find few transcriptomic states reflective of distinct cell types associated with the large epithelial partitions, but rather separations that are reflective of maturation of these tissues from the embryonic state (18–48 hours) into the adult state that is maintained after embryogenesis has completed. While we can fully recapitulate the annotations provide in our previous works (1,3), we extend our observations of the data to include further characterization of the putative stem cell population, including providing a full series of sperm maturation that has yet to be fully analysed, and the identification of two separate immune-like profiles associated with both germ layers. We hope that the updated catalog of transcriptomic states provided here will serve the community in its quest to understand not only *Nematostella* biology, but also as a valuable resource for use in comparative analyses of cell type composition across evolutionary time.

## Methods

### Additional data:

Cell suspensions were generated according to Steger for an additional 24hr gastrula sample, an additional 2d planula, and an additional mesentery sample that was harvested from an adult male after a regular spawning cycle. Nonetheless, the mesentery was washed multiple times to reduce the amount of

sperm present in the capture. The 24hr and mesentery samples were dissociated with collagenase, whereas the 2d planula samples were processed with tripleE. Raw data from these cell captures have been deposited into the GEO repository (GSE200198).

## Mapping:

Raw sequence data from all samples (GSE200198 and GSE154105) were mapped to the Nv2 gene models and accompanying vs.2 chromosome genome assembly ((4) [https://simrbase.stowers.org/jb\\_pub/?data=data/starlet\\_pub](https://simrbase.stowers.org/jb_pub/?data=data/starlet_pub)) using the 10x genomics cell ranger pipeline, without secondary analysis and force recovery of 10,000 cells per library.

## Single cell analyses

The UMI-reduced gene by cell matrices were then imported into R and filtered for cells containing at least 250 UMIs. Putative multiplets were identified as outliers at the upper end of the UMI content and removed for each sample prior to merging all data together. Analysis and clustering strategy follows that reported in (1), and the R script can be accessed on our GitHub ([https://github.com/technau/Nv2\\_Atlas](https://github.com/technau/Nv2_Atlas)). Briefly, expression matrices from each sample were merged, the data was then log normalized (Seurat::NormalizeData), the collection of top 1000 variable features from each sample were scaled within each sample separately (split.by parameter), which was used as input into the Seurat::RunPCA function, and finally 30 dimensions were used as input for both Seurat::FindNeighbors and Seurat::RunUMAP. The entire dataset was then clustered into 12 major cell-type partitions (resolution = 0.2) and each partition was then re-analyzed separately to identify transcriptomic profiles (clusters) with robust unique gene sets. The top 2000 variable gene set from the partition was identified (Seurat::FindVariableFeatures), these were scaled and centered by sample across the entire sample dataset (Seurat::ScaleData, split.by='orig.ident') and imported into the scale.data slot of the partition. PCA and clustering was then performed on all partition subsets separately. Separate analyses of the tissue layers were similarly performed on data subsets containing multiple partitions. Semi-automated cluster naming for each partition was performed with the marker gene sets found in **Supplemental Data (2)**. UMAP reduction cell plots with underlying expression matrix will be available for exploration on the UCSC Cellbrowser website ([sea-anemone-atlas.cells.ucsc.edu](http://sea-anemone-atlas.cells.ucsc.edu))

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors have approved the submission for publication

### Availability of data and material

The datasets generated and analysed during the current study are available in the Gene expression Omnibus (GEO) repository, <https://www.ncbi.nlm.nih.gov/geo>

## Competing interests

Authors declare no competing interests.

## Funding

BW was funded by NIH/NIMH RF1MH132662, CIRM DISC0-14514 and NIH/NHGRI U24HG002371. This work was funded by grants from the Austrian Science Fund FWF P27353 (UT) and P31018 (AGC).

## Author Contributions

JS, JH, AD, PK, PM, SN, and AGC generated sc transcriptome datasets used in this work. JDM generated the mapping tool for generating the expression matrices; BW imported and formatted the data for hosting on the UCSC Cellbrowser website; AGC re-analyzed all data. AGC and UT wrote the manuscript.

## Acknowledgements

The authors thank the community for continued feedback regarding the utility of the dataset.

## References

1. 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.
2. Seb e-Pedr os A, Saudemont B, Chomsky E, Plessier F, Mailh e 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.
3. 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.
4. Zimmermann B, Montenegro JD, Robb SMC, et al. Topological structures and syntenic conservation in sea anemone genomes [2023]. *Nat Commun.* 14:8270 doi/10.1038/s41467-023-44080-7
5. Kraus Y, Aman A, Technau U, Genikhovich G. Pre-bilaterian origin of the blastoporal axial organizer. *Nat Commun.* 2016 May 27;7(1):11694.
6. Lecl ere L, Bause M, Sinigaglia C, Steger J, Rentzsch F. Development of the aboral domain in *Nematostella* requires  $\beta$ -catenin and the opposing activities of *six3/6* and *frizzled5/8*. *Development.* 2016 Jan 1;dev.120931.
7. Lebedeva T, Aman AJ, Graf T, Niedermoser I, Zimmermann B, Kraus Y, et al. Cnidarian-bilaterian comparison reveals the ancestral regulatory logic of the  $\beta$ -catenin dependent axial patterning. *Nat Commun.* 2021 Jun 29;12(1):4032.

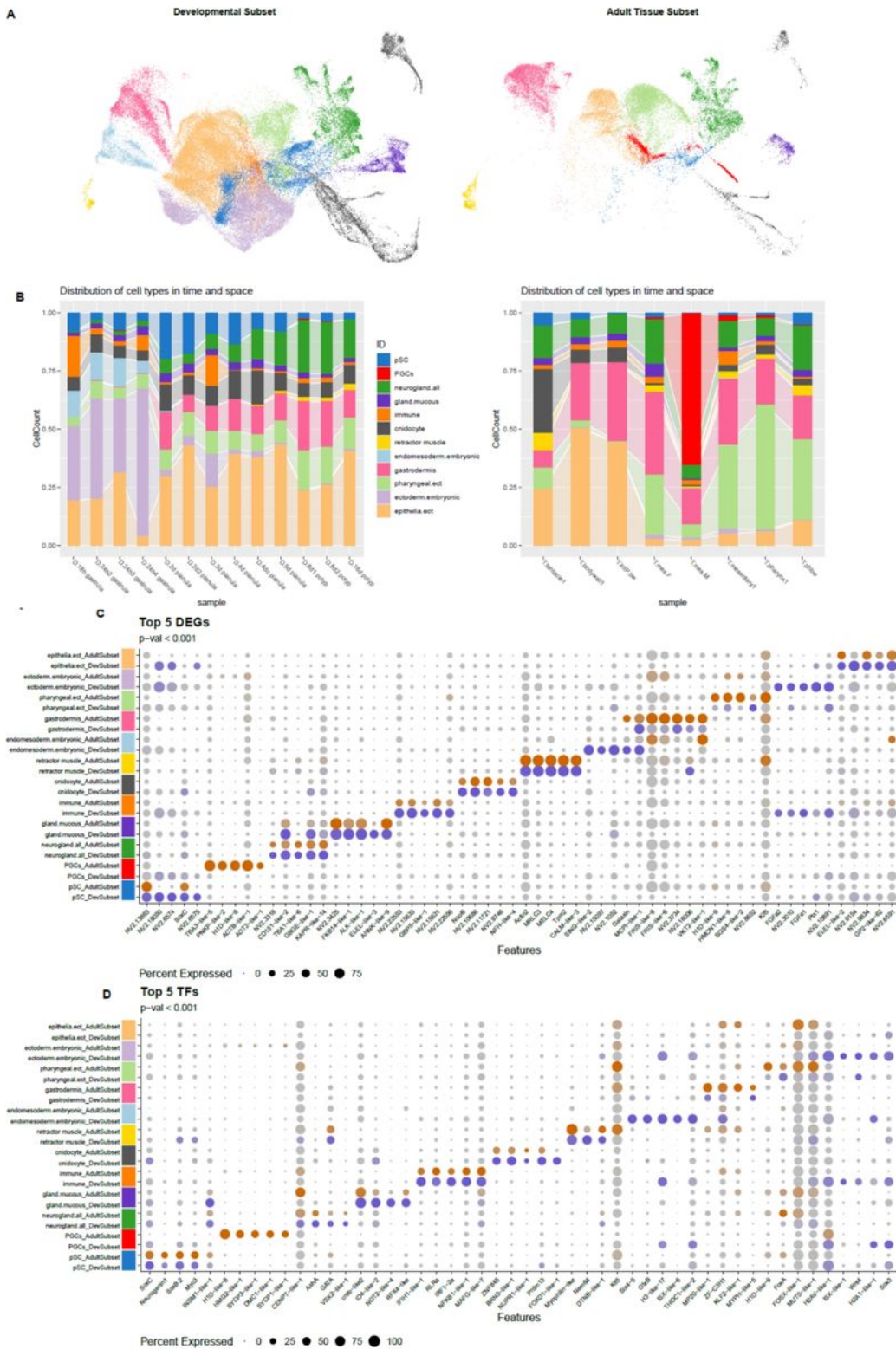
8. Fritzenwanker JH, Saina M, Technau U. Analysis of forkhead and snail expression reveals epithelial–mesenchymal transitions during embryonic and larval development of *Nematostella vectensis*. *Dev Biol*. 2004 Nov;275(2):389–402.
9. Kusserow A, Pang K, Sturm C, Hroudá M, Lentfer J, Schmidt HA, et al. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*. 2005 Jan 13;433(7022):156–60.
10. Sinigaglia C, Busengdal H, Leclère L, Technau U, Rentzsch F. The Bilaterian Head Patterning Gene *six3/6* Controls Aboral Domain Development in a Cnidarian. Levine M, editor. *PLoS Biol*. 2013 Feb 19;11(2):e1001488.
11. Columbus-Shenkar YY, Sachkova MY, Macrander J, Fridrich A, Modepalli V, Reitzel AM, et al. Dynamics of venom composition across a complex life cycle. *eLife*. 2018 Feb 9;7:e35014.
12. Orts DJB, Moran Y, Cologna CT, Peigneur S, Madio B, Praher D, et al. BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker. *FEBS J*. 2013 Oct;280(19):4839–52.
13. Rentzsch F, Fritzenwanker JH, Scholz CB, Technau U. FGF signalling controls formation of the apical sensory organ in the cnidarian *Nematostella vectensis*. *Development*. 2008 May 15;135(10):1761–9.
14. Sinigaglia C, Busengdal H, Lerner A, Oliveri P, Rentzsch F. Molecular characterization of the apical organ of the anthozoan *Nematostella vectensis*. *Dev Biol*. 2015 Feb;398(1):120–33.
15. Gilbert E, Teeling C, Lebedeva T, Pedersen S, Christmas N, Genikhovich G, et al. Molecular and cellular architecture of the larval sensory organ in the cnidarian *Nematostella vectensis*. *Development*. 2022 Aug 15;149(16):dev200833.
16. Steinmetz PRH, Aman A, Kraus JEM, Technau U. Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. *Nat Ecol Evol*. 2017 Sep 11;1(10):1535–42.
17. Burtscher I, Lickert H. *Foxa2* regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development*. 2009 Mar 15;136(6):1029–38.
18. Wang P, McKnight KD, Wong DJ, Rodriguez RT, Sugiyama T, Gu X, et al. A Molecular Signature for Purified Definitive Endoderm Guides Differentiation and Isolation of Endoderm from Mouse and Human Embryonic Stem Cells. *Stem Cells Dev*. 2012 Aug 10;21(12):2273–87.
19. Koo BK, Spit M, Jordens I, Low TY, Stange DE, Van De Wetering M, et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature*. 2012 Aug;488(7413):665–9.
20. Hao HX, Xie Y, Zhang Y, Charlat O, Oster E, Avello M, et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature*. 2012 May;485(7397):195–200.
21. Ikmi A, Steenbergen PJ, Anzo M, McMullen MR, Stokkermans A, Ellington LR, et al. Feeding-dependent tentacle development in the sea anemone *Nematostella vectensis*. *Nat Commun*. 2020 Sep 2;11(1):4399.
22. Fritz AE, Ikmi A, Seidel C, Paulson A, Gibson MC. Mechanisms of tentacle morphogenesis in the sea anemone *Nematostella vectensis*. *Development*. 2013 May 15;140(10):2212–23.
23. Trevino M, Stefanik DJ, Rodriguez R, Harmon S, Burton PM. Induction of canonical Wnt signaling by alsterpaullone is sufficient for oral tissue fate during regeneration and embryogenesis in

- Nematostella vectensis. Dev Dyn. 2011 Dec;240(12):2673–9.
24. Schaffer AA, Bazarsky M, Levy K, Chalifa-Caspi V, Gat U. A transcriptional time-course analysis of oral vs. aboral whole-body regeneration in the Sea anemone *Nematostella vectensis*. BMC Genomics. 2016 Dec;17(1):718.
  25. Chen CY, McKinney SA, Ellington LR, Gibson MC. Hedgehog signaling is required for endomesodermal patterning and germ cell development in the sea anemone *Nematostella vectensis*. eLife. 2020 Sep 24;9:e54573.
  26. He S, Shao W, Chen S (Cynthia), Wang T, Gibson MC. Spatial transcriptomics reveals a cnidarian segment polarity program in *Nematostella vectensis*. Curr Biol. 2023 Jul;33(13):2678–2689.e5.
  27. Marlow HQ, Srivastava M, Matus DQ, Rokhsar D, Martindale MQ. Anatomy and development of the nervous system of *Nematostella vectensis*, an anthozoan cnidarian. Dev Neurobiol. 2009 Mar;69(4):235–54.
  28. 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.
  29. Rentzsch F, Layden M, Manuel M. The cellular and molecular basis of cnidarian neurogenesis. WIREs Dev Biol. 2017 Jan;6(1):e257.
  30. 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.
  31. Lemaître QIB, Bartsch N, Kouzel IU, Busengdal H, Richards GS, Steinmetz PRH, et al. NvPrdm14d-expressing neural progenitor cells contribute to non-ectodermal neurogenesis in *Nematostella vectensis*. Nat Commun. 2023 Aug 10;14(1):4854.
  32. Tournière O, Gahan JM, Busengdal H, Bartsch N, Rentzsch F. *Insm1*-expressing neurons and secretory cells develop from a common pool of progenitors in the sea anemone *Nematostella vectensis*. Sci Adv. 2022 Apr 22;8(16):eabi7109.
  33. Busengdal H, Rentzsch F. Unipotent progenitors contribute to the generation of sensory cell types in the nervous system of the cnidarian *Nematostella vectensis*. Dev Biol. 2017 Nov;431(1):59–68.
  34. Lang T, Hansson GC, Samuelsson T. Gel-forming mucins appeared early in metazoan evolution. Proc Natl Acad Sci. 2007 Oct 9;104(41):16209–14.
  35. Margolis SR, Dietzen PA, Hayes BM, Wilson SC, Remick BC, Chou S, et al. The cyclic dinucleotide 2'3'-cGAMP induces a broad antibacterial and antiviral response in the sea anemone *Nematostella vectensis*. Proc Natl Acad Sci. 2021 Dec 21;118(51):e2109022118.
  36. Lewandowska M, Sharoni T, Admoni Y, Aharoni R, Moran Y. Functional Characterization of the Cnidarian Antiviral Immune Response Reveals Ancestral Complexity. Whittington C, editor. Mol Biol Evol. 2021 Sep 27;38(10):4546–61.
  37. Wolenski FS, Garbati MR, Lubinski TJ, Traylor-Knowles N, Dresselhaus E, Stefanik DJ, et al. Characterization of the Core Elements of the NF- $\kappa$ B Signaling Pathway of the Sea Anemone

*Nematostella vectensis*. Mol Cell Biol. 2011 Mar 1;31(5):1076–87.

38. Wolenski FS, Bradham CA, Finnerty JR, Gilmore TD. NF- $\kappa$ B is required for cnidocyte development in the sea anemone *Nematostella vectensis*. Dev Biol. 2013 Jan;373(1):205–15.
39. Brennan JJ, Messerschmidt JL, Williams LM, Matthews BJ, Reynoso M, Gilmore TD. Sea anemone model has a single Toll-like receptor that can function in pathogen detection, NF- $\kappa$ B signal transduction, and development. Proc Natl Acad Sci [Internet]. 2017 Nov 21 [cited 2023 Sep 19];114(47). Available from: <https://pnas.org/doi/full/10.1073/pnas.1711530114>
40. Snyder GA, Eliachar S, Connelly MT, Talice S, Hadad U, Gershoni-Yahalom O, et al. Functional Characterization of Hexacorallia Phagocytic Cells. Front Immunol. 2021 Jul 26;12:662803.
41. Levy S, Elek A, Grau-Bové X, Menéndez-Bravo S, Iglesias M, Tanay A, et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. Cell. 2021 May;184(11):2973–2987.e18.

## Figures



**Figure 1**

Updated cell atlas for the starlet sea anemone *Nematostella vectensis*. **A)** Dimensional reduction cell map (UMAP) illustrating clustering of cells from the developmental (left) or adult tissue (right) subsets. **B)** Relative contribution of each library to the clusters from the developmental (left) or tissue (right) samples. **C,D)** Dotplot representation of differentially expressed gene sets from all genes (C) and restricted to



transcription factors only (D). Expression profiles are split between the two subsets. Orange scale = Adult subset; Blue scale = Developmental subset.

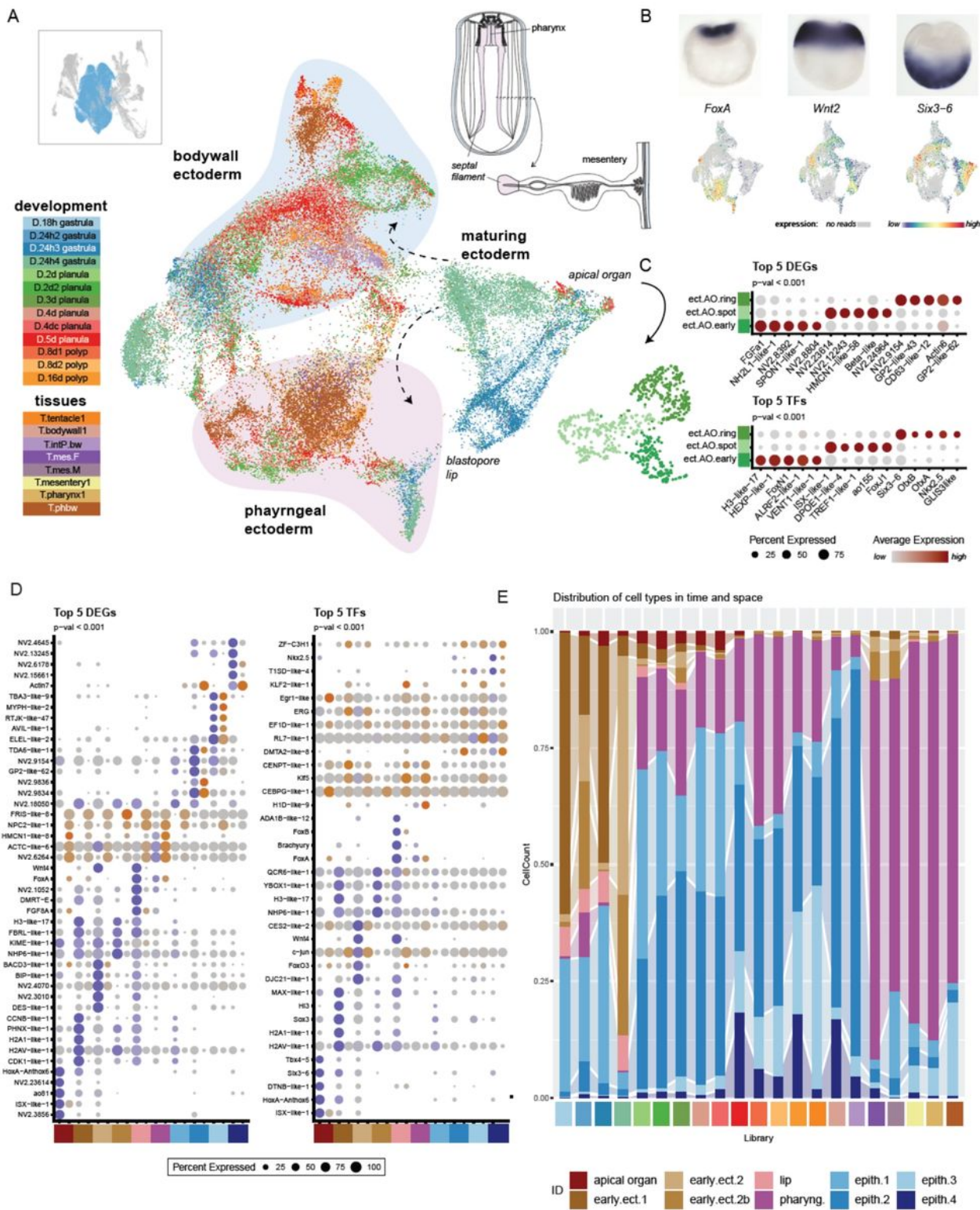
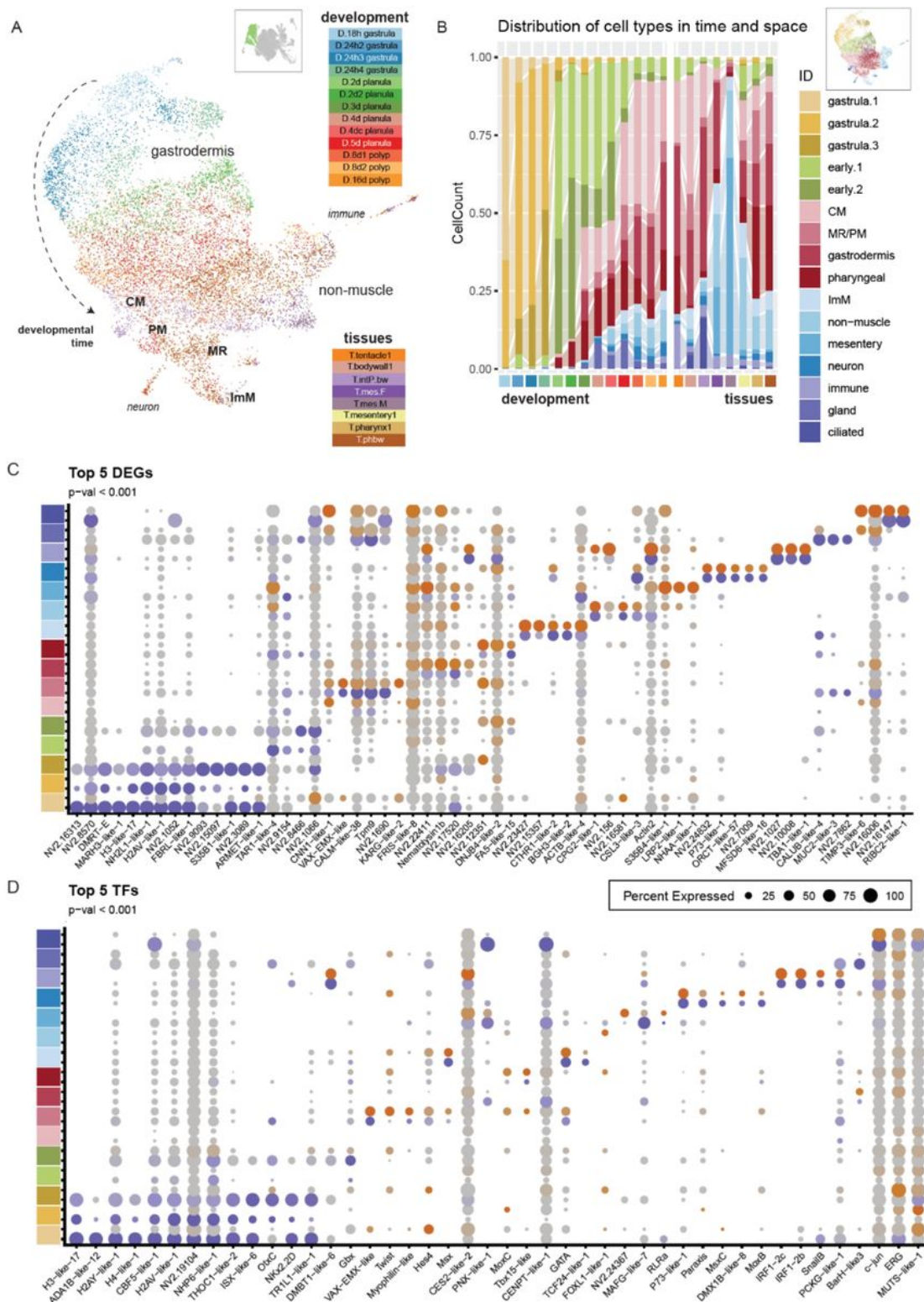


Figure 2

**Early embryonic ectoderm matures into two distinct tissues.** A) UMAP dimensional reduction of the ectodermal partition colored by sample origin. The early transcriptomic state sorts along the oral – aboral

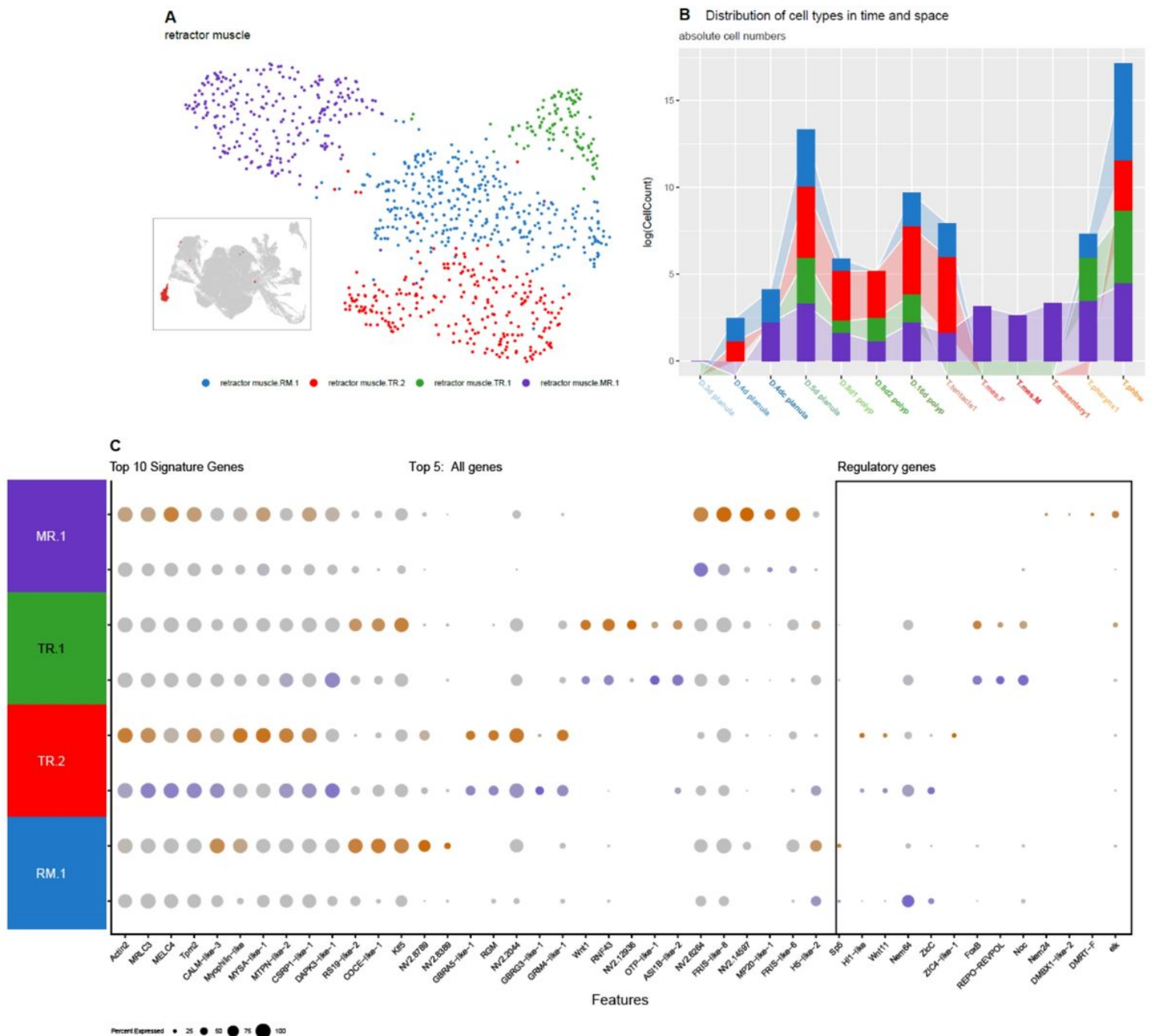
axis. Similarly, the mature transcriptomic state separates into the orally derived pharyngeal ectoderm of the pharynx and septal filaments and the aborally-derived outer epithelial layer. **B)** Sorting according to ectodermal patterning within the transcriptomic data is evidenced by comparing expression patterns known from in situ hybridization (top) with expression in the dimensional reduction (bottom). **C)** UMAP (left) and Dotplot of differentially expressed genes (top) and transcription factors (bottom) of the apical organ cells. **D)** Dotplot indicating expression of differentially expressed genes (left) and transcription factors (right) across all ectodermal 10 clusters. Expression separated between cells of the developmental series (blue scale) and the adult tissue series (orange scale). Grey indicates average scaled expression of 0 or below. **E)** Relative distribution of cell cluster identities across all samples in the dataset.



**Figure 3**

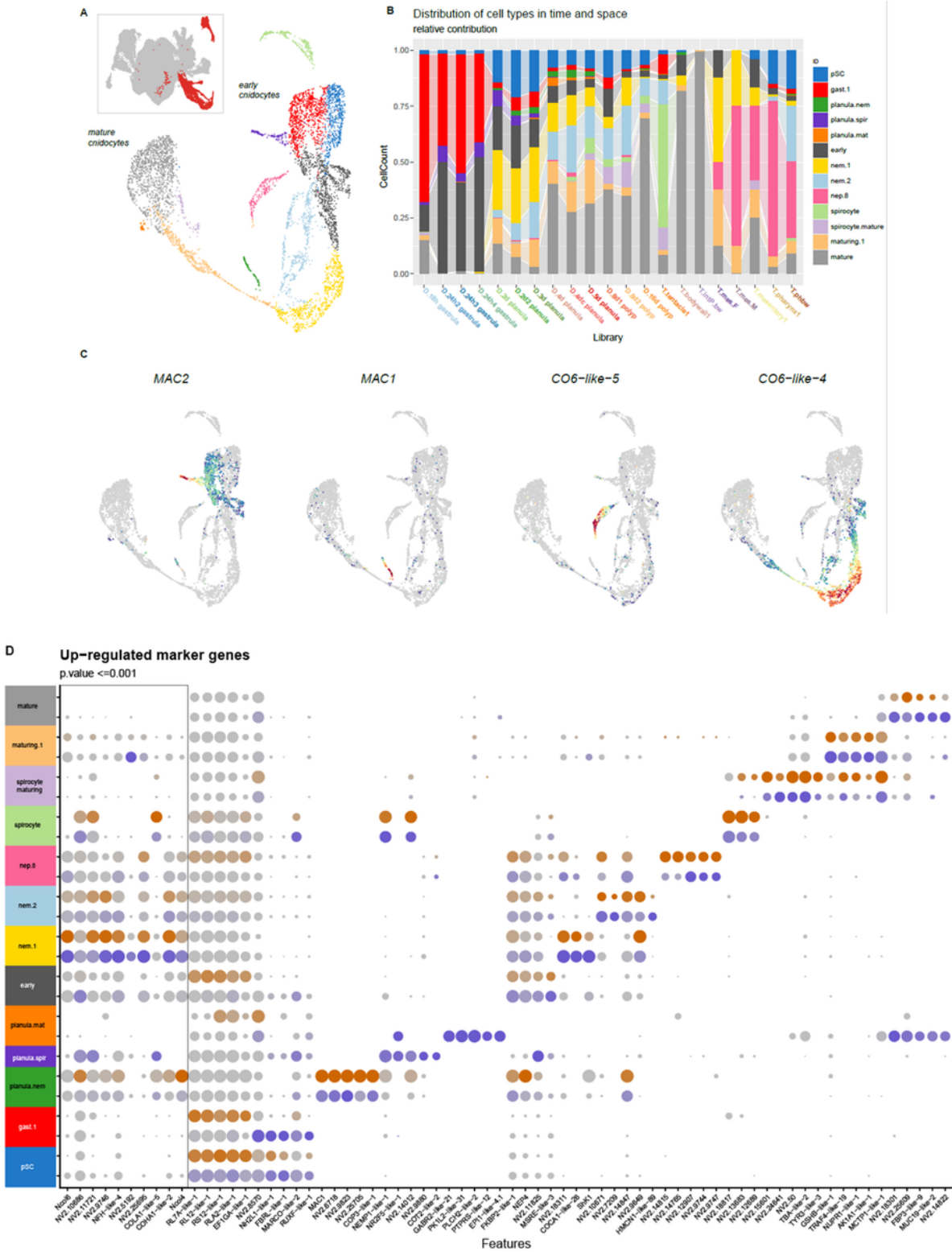
**Inner cell layer matures into gastrodermis and derived cell types.** **A)** UMAP dimensional reduction cell plot, colored by sample origin. Inset: included partitions. Dashed arrow: maturation trajectory. CM: circular muscle; PM: parietal muscle; MR: mesentery retractor muscle; ImM: intermuscular membrane. **B)** Bar plot representation of fraction of cells from each cluster (colours) within each sample (bars). **C,D)** Dot plot representation of expression profiles of up-regulated genes (C) and up-regulated transcription factors (D)

across each cluster. Expression separated between cells of the developmental series (blue scale) and the adult tissue series (orange scale). Grey indicates average scaled expression of 0 or below.



**Figure 4**

**Sub-states of retractor muscle cells from both germ layers can be resolved.** **A)** Retractor muscle partition within the entire dataset (inset). UMAP representation of four unique cluster states. **B)** Bar plot showing the log count (C) of cells from each sample contributing to the partition. **C)** Dot plot of gene expression separated between partition cluster cells from the developmental (blue gradient) or adult tissue (orange gradient) samples. Expression profiles of the differentially expressed genes across the partition (Signature Genes) and each partition cluster (Top 5 All Genes), and its regulatory profile (box).



**Figure 5**

**Cnidocyte specification pathways are recovered.** **A** Cell plot indicating the trajectory clusters of the cnidocyte partition (inset). **B** Distribution of sample contribution to each identified cluster. **C** Specific toxin profiles associated with cnidocyte subtypes. **D** Dotplot of marker expression (box) and differentially up-regulated genes of each cluster. See Supplementary material for full gene lists.

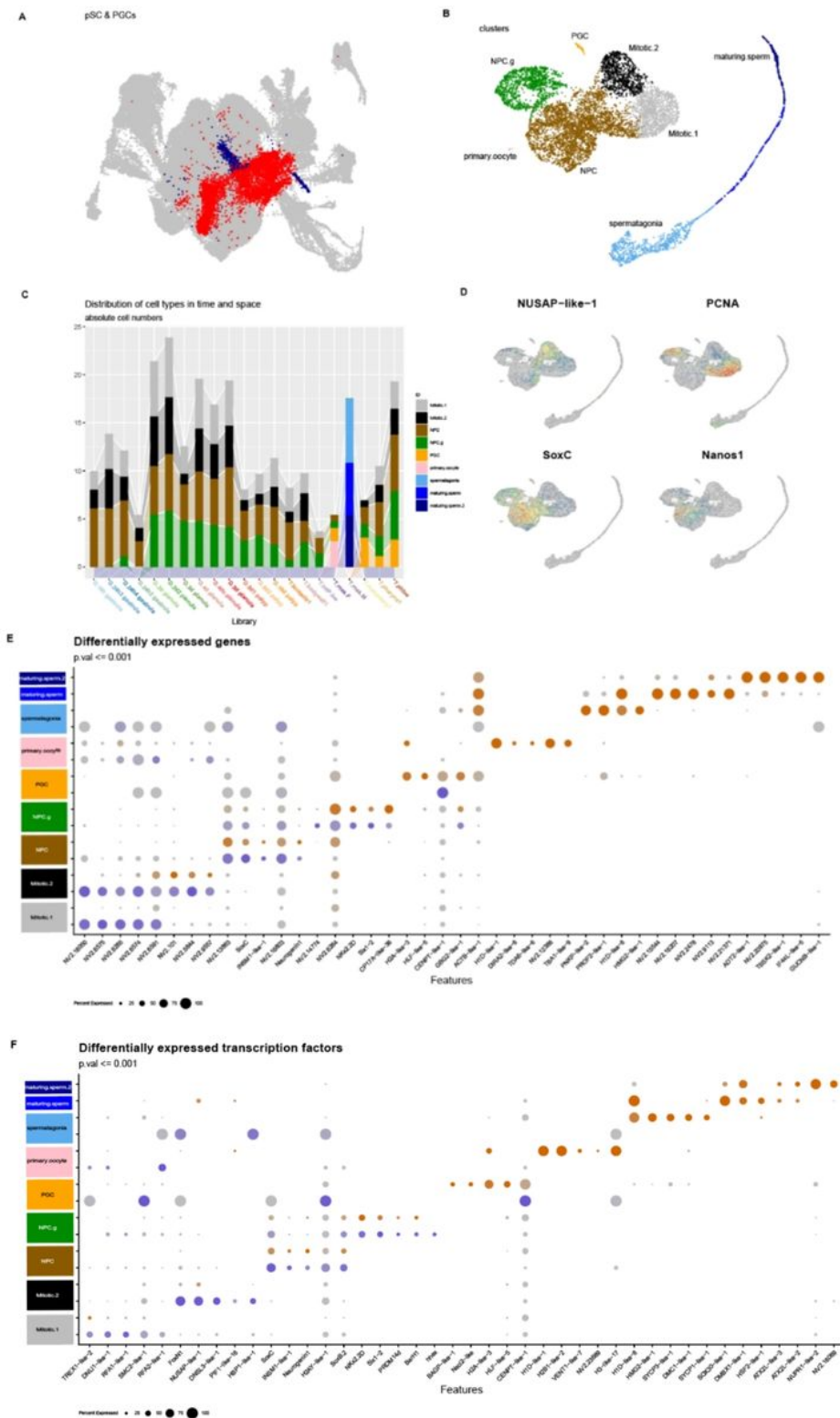


Figure 6

Neuroglandular precursor cells can be identified exiting from the mitotic cycle and derive from both germ layers. **A**. Partitions included in the analyses: Putative stem cells (red) and primary germ cells (blue) **B** Identified clusters. **C** Distribution of cells from each sample included in the cluster: absolute cell counts **D** expression profiles of mitotic markers of DNA-synthesis (*PCNA*) and division (*NUSAP-like-1*), and NPC

marker *soxC* and pre-neural marker *nanos1*. **E,F** Dotplot expression of top marker genes (E) and differentially expressed transcription factors (F) from each cluster.

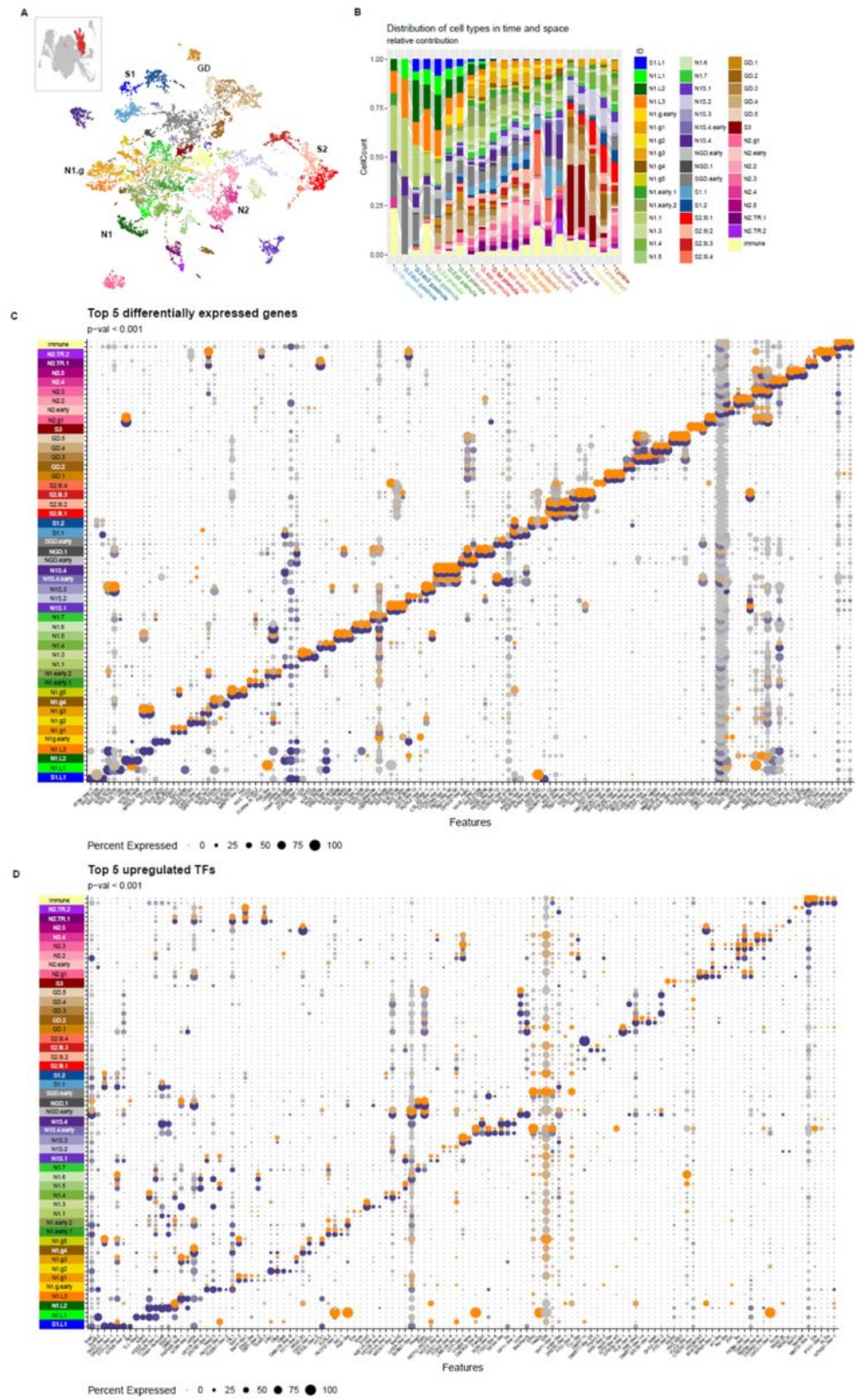
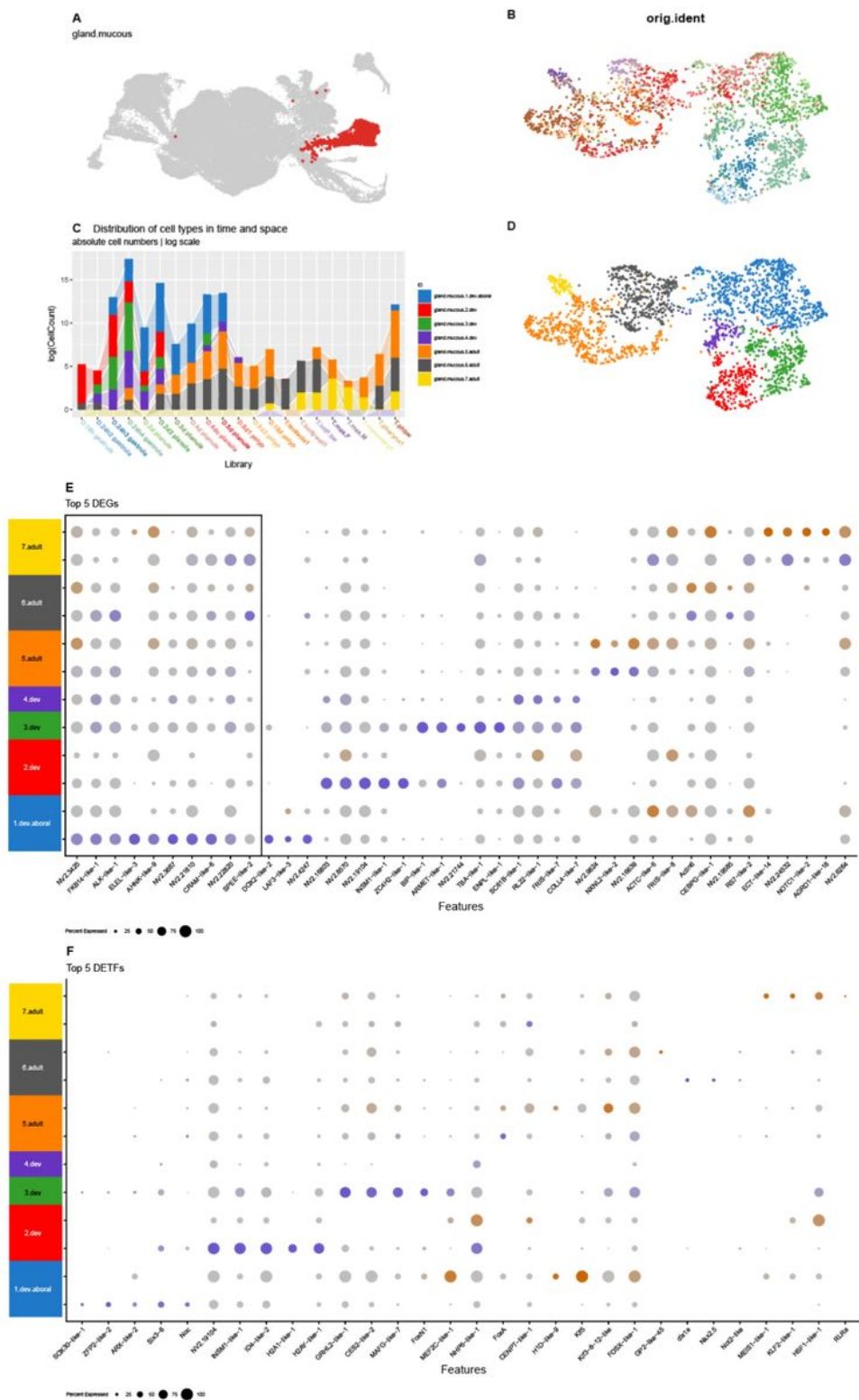


Figure 7

Neuroglular derivatives include cells of multiple related cell types. **A** Dimensional reduction cell plot (UMAP) showing clustering of the neuroglular partition (inset). **B** Bar plot of relative contribution of

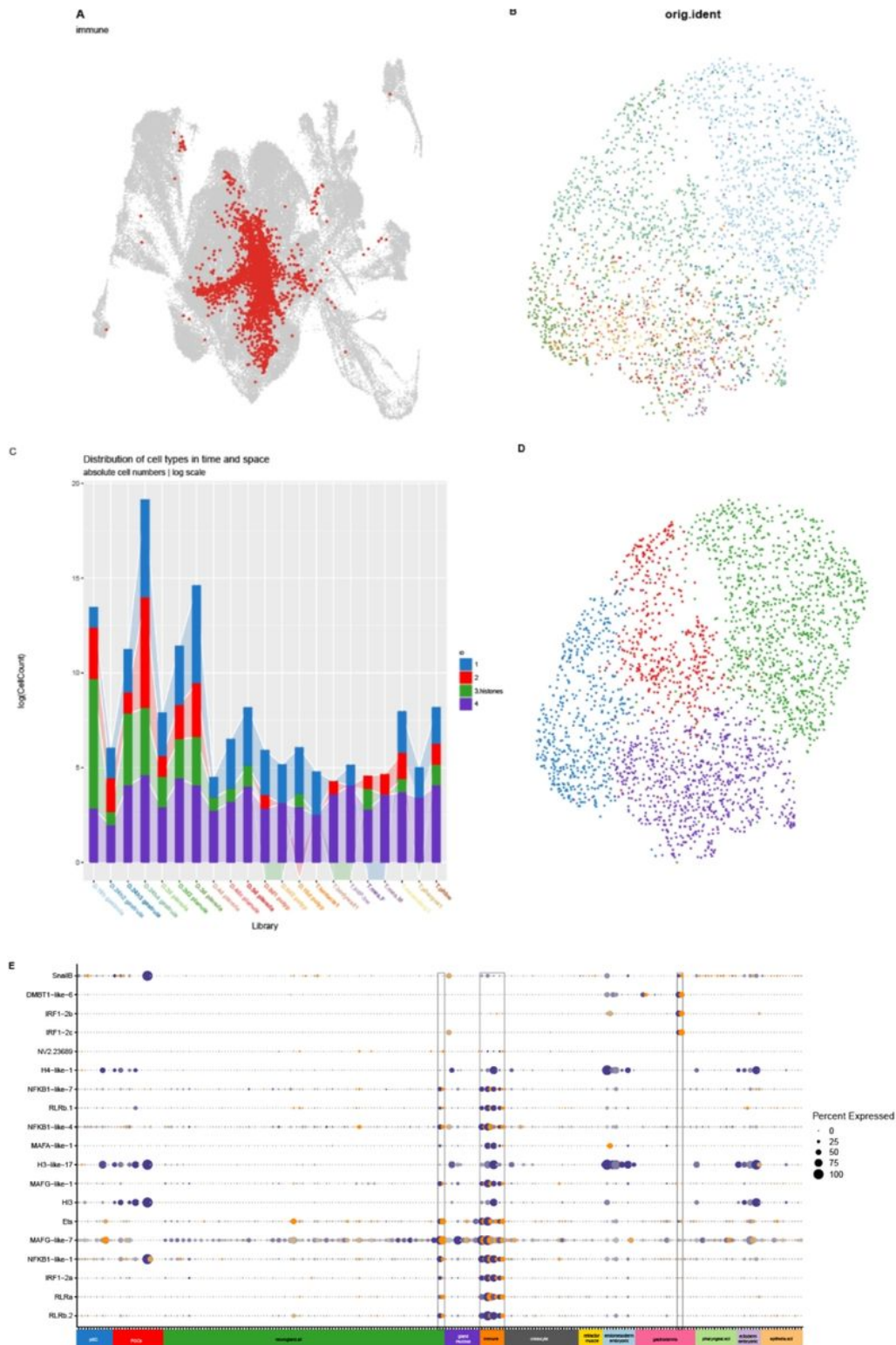
each cell state across all samples. **C,D** Dot plot of gene expression of the top 5 state-specific gene sets from all genes (C) and the set of upregulated transcription factors (D).



**Figure 8**

Mucin gland cells are homogeneous. Figure composition as in figure 4. Differentially regulated transcription factors are not shown but can be found in the accompanying supplementary gene lists.





**Figure 9**

**Cells expressing an immune response are present across all samples. A** immune partition within the entire dataset. **B,C** Cell plots (UMAP) coloured by sample (B) and cell states (C) within the partition. **D** Bar plots showing the distribution of cells across all samples in terms of absolute number of cells (log values). **E** Dotplot expression profile of specific immune related regulatory genes across the entire dataset. The signature is found within the immune partition (orange) but also in the immune-cells of the

neuroglandular partition (box in green partition), but not shared with the putative immune signature of the inner cell layer (box in pink partition).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalData1.xlsx](#)
- [SupplementalData2.xlsx](#)
- [SupplementalFigures.pdf](#)