A single-cell atlas of the free-living miracidium larva of *Schistosoma mansoni*

Teresa Attenborough[1,3]*, Kate A. Rawlinson[1,2]*, Carmen L. Diaz Soria[1], Kirsty Ambridge[1], Geetha Sankaranarayanan[1], Jennie Graham[1], Stephen R. Doyle[1], Gabriel Rinaldi[1,4], Matthew Berriman[1,3] (*These authors contributed equally).

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB101SA, UK
2. Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Massachusetts, USA.
3. Wellcome Centre for Integrative Parasitology, Institute of Infection, Immunity & Inflammation, College of Medical,Veterinary & Life Sciences, University of Glasgow, 120 University Place, Glasgow G12 8TA, UK
4. Department of Life Sciences, Aberystwyth University, Edward Llwyd Building, Penglais Campus, Aberystwyth SY23 3DA, UK

Abstract

Schistosomes are parasitic flatworms whose infections affect millions of people worldwide. The parasite has a complex life cycle that is poorly understood at a cellular level. We have developed a cellular transcriptome atlas of the miracidium larva of *Schistosoma mansoni,* the first post-embryonic stage of development. The miracidium infects a snail host, inside which its stem cells generate daughter sporocysts that give rise to human-infective cercariae larvae. We use a combination of single-cell RNA sequencing, *in situ* hybridisation and image analysis to characterise the miracidia. Each larva comprises ~365 cells and 19 transcriptionally distinct cell types. We show that 93% of miracidium cells are somatic (>45% neural, 19% muscle, 13% tegument, 2% parenchyma, 2% protonephridia) and the remaining 7% are stem cells. Cellular diversity within tissue types is revealed, including two transcriptionally distinct muscle clusters, corresponding to the circular and the longitudinal muscles of the body wall, as well as two stem cell populations that show different activation states and cluster further by sex. This whole-body atlas provides new insight into the organisation of the larva and lays the foundation for understanding how the somatic cells drive the behaviour and physiology necessary for infection and propagation of the stem cells inside the snail.

Introduction

The parasitic blood fluke *Schistosoma mansoni* is a causative agent of the neglected tropical disease schistosomiasis, that causes up to 200,000 human deaths/annum (McManus et al; 2018). Enabled by the advent of single-cell RNA sequencing (scRNA-seq), molecular cell-type classification has been initiated for this trematode flatworm (Wang et al., 2018; Tarashansky et al., 2019; Li et al., 2020; Diaz et al., 2020; Wendt et al., 2020). These scRNA-seq studies have opened the door to investigate schistosome developmental cell biology and to identify new targets for controlling infection and transmission of the disease. To date scRNA-seq studies have identified heterogeneity among stem cells that drive schistosome development and reproduction, classified a variety of differentiated cell types across the life cycle and have revealed several key genes and their native cell types that are essential for parasite survival and propagation (reviewed in Sarfati et al 2021).

The life cycle of *Schistosoma mansoni*, like that of other schistosome species, is complex, involving parasitic stages in two hosts (a snail and a mammal), with short-lived, freshwater larval stages in between. Mature male and female worms reside in the portal veins of the mammalian host, and fertilisation of the zygote occurs inside the female worm (Jurberg et al., 2009). Embryogenesis proceeds as the egg is laid into the host, and over the ensuing six days (Ashton et al., 2001) passes through the host's endothelium into the intestine and is passed in faeces into the outside environment (Costain et al., 2018). If the egg lands in fresh water a miracidium larva hatches out, swims and infects a snail host. Once inside the snail the miracidium becomes known as the mother sporocyst, and the stem (or germinal) cells produce many daughter sporocysts that migrate within the snail. In some of these daughter sporocysts a second round of embryogenesis generates the cercariae larvae asexually from a stem cell. The cercariae emerge from the snail and infect mammals/humans. Inside the mammal, the cercariae undergoes metamorphosis to the schistosomula stage, discarding its tail and remodelling the tegument/skin of the rest of the body. The schistosomulum then migrates into the blood vessels and over the course of the following 5 weeks grows into a juvenile stage that travels from the lungs to the liver and to the hepatic portal vein where male and female worms pair, mature and mate. Single-cell RNAseq studies for the whole animal have only been carried out on intra-mammalian stages to date; the schistosomulum (Diaz et al., 2020), juvenile (3.5 weeks post-infection of mouse, Li et al., 2020) and adult (Wendt et al., 2020), whereas single cell studies on stem cells only have been undertaken in

the mother sporocyst (Wang et al., 2018) and juvenile stages (2.5 and 3.5 weeks post infection)(Tarashansky et al., 2019).

To study schistosome development, scRNA-seq experiments can facilitate the reconstruction of cell type developmental trajectories *in silico* if the temporal resolution is fine enough. To add a new developmental time point to existing information, and to produce a foundation for all existing and subsequent scRNAseq studies, we have generated a cell type transcriptome atlas for the earliest accessible developmental stage, the miracidium larva. The six days of embryonic development leading to hatching of the miracidium (Ashton et al., 2001) give rise to a rich set of cell types that have been identified at the ultrastructural level using transmission electron microscopy (Pan, 1980). These cells have classically been organised into eight major systems: epithelial, terebratorium, musculature, interstitial cells, penetration glands, excretory, nervous, and stem/germinal cells (Pan, 1980). The role of the miracidium, and its differentiated somatic cells, is to carry the stem cells to a favourable environment inside the snail where they can clonally proliferate to generate daughter sporocysts, and subsequently cercariae, to propagate the life cycle. The stem cells in the miracidia are thought to be quiescent (Pan, 1980; Wang et al, 2013) and start dividing 20 hours post-infection (Wang et al., 2013). However, the miracidium is a non-feeding larva (lacking a mouth and gut, Pan, 1980) and relies on internal stores of energy to sustain swimming, and to locate and infect its snail host; if it fails to do this within 12 hours it dies (Maldonado and Matienzo, 1948; Chernin, 1968). To enable investigation into the function, regulation and interactions of these cell types, and to understand how the larva optimises its chance of encountering and infecting a snail in such a short timeframe, we have created a reference cell type transcriptome atlas that provides molecular definitions of the cell types described by Pan (1980).

Results

The miracidium has 365 nuclei

To estimate the number of cells in a miracidium, we traced and counted every nucleus from confocal z-stacks of DAPI-stained individuals using the TrakEM2 plugin in ImageJ (Cardona et al., 2012). There are approximately 365 (n = 3) nuclei in a miracidium. However, because some cell types are multinucleate, such as the apical gland and epithelial cells, and others are anucleate, e.g. the ciliary plates (Pan, 1980), this number provides an estimation of the true number of cells (Fig. 1A).

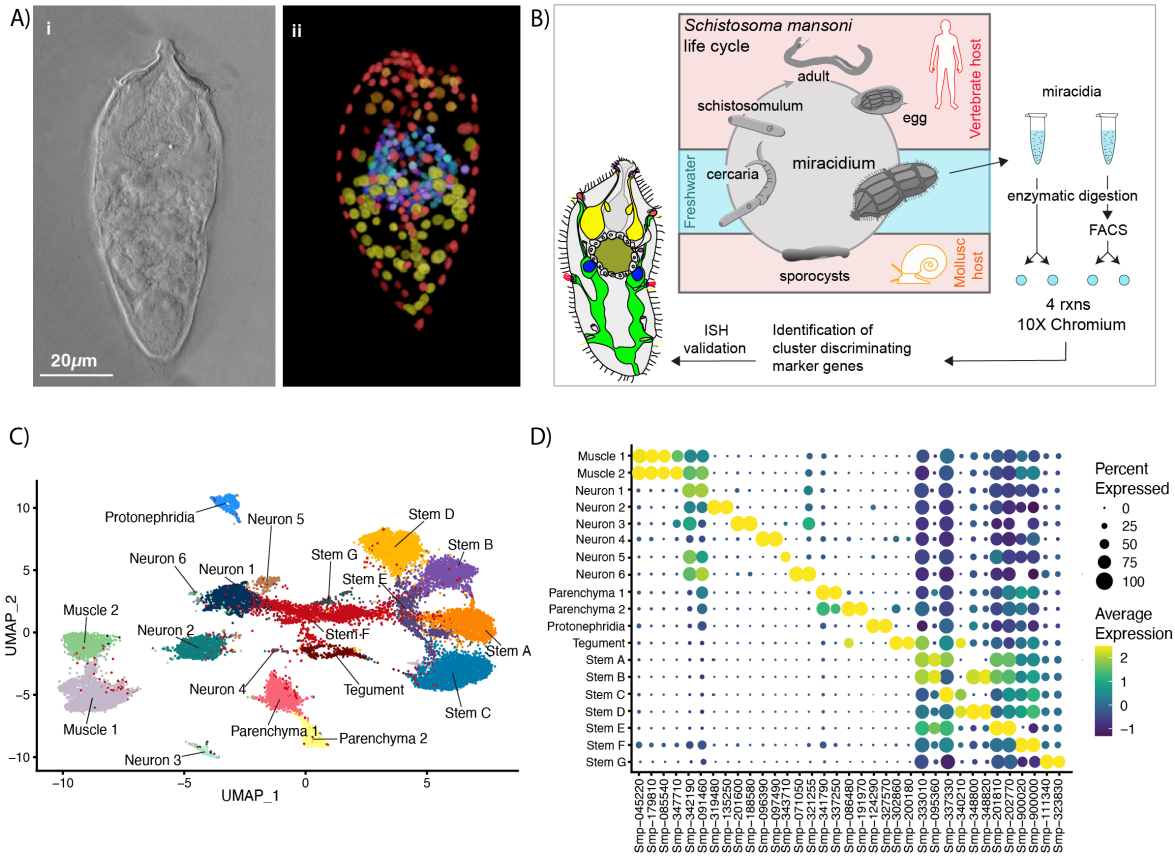


Figure 1. Identification of 19 transcriptionally distinct cell types in the miracidium
A) The miracidium is composed of ~365 cells; **i)** DIC image of miracidium, **ii)** 3D projection of confocal z-stack of DAPI-stained miracidium with nuclei segmented to enable counting (anterior top in all images). **B)** Experimental diagram describing the *Schistosoma mansoni* life cycle, single-cell RNAseq analysis of the miracidium and validation pipeline. An average of 9,975 miracidia per sample were dissociated; one sample was enriched for live cells (propidium iodide negative) using fluorescence-activated cell sorting (FACS), and the other sample was left unsorted. Cells were loaded into 4 10X reactions; two sorted and two unsorted, according to the 10X Chromium single-cell 3' protocol. Clustering was carried out to identify distinct populations and population-specific markers. Validation of population-specific markers was performed by *in situ* hybridisation (ISH). **C)** Uniform Manifold Approximation and Projection (UMAP) representation of 20,478 miracidium single cells. Cell clusters are coloured and distinctively labelled (resolution 1 and 55 PCs). **D)** Gene-expression profiles of top two population markers identified for each of the cell clusters. The colours represent the level of expression from yellow (high expression) to dark blue (low expression). The sizes of the circles represent the percentages of cells in those clusters that expressed a specific gene.

scRNA-seq of 20,478 miracidia cells

We performed scRNAseq on a pool of mixed-sex miracidia, collected within four hours of hatching from eggs (Fig. 1B) A novel protocol was developed to efficiently dissociate the

parasites using a protease cocktail. Following dissociation, four samples of the cell suspensions were collected; two of which were enriched for live cells using fluorescence-activated cell sorting (FACS) and two were unsorted (Fig. 1B). Using the droplet-based 10X Genomics Chromium platform, transcriptome-sequencing data were generated from a total of 33,391 cells, of which 20,478 passed strict quality-control filters, resulting in a median of 858 genes and 1,761 median UMI counts per cell (Supplementary Table 1). Given that an individual miracidium comprises 365 nuclei (Fig. 1A), the number of quality-controlled cells theoretically represents >56× coverage of each cell in the larva.

Using Seurat, 19 distinct clusters of cells were identified, along with putative marker genes best able to discriminate between the populations (Fig. 1C & D and Supplementary Table 2 and 3). Manually curated lists of previously defined cell-specific genes, from later developmental stages (Diaz et al 2020; Wendt et al 2020; Sarfati et al 2021), were compared against the list of putative markers and *in silico* gene expression was examined to identify the cell types that each Seurat cluster represented. Based on these markers, the 19 clusters were resolved into the following populations of cells: two muscle-like (3,934 cells), six resembling neurons (3,255 cells), one tegumental (366 cells), two parenchymal (1698 cells), one protonephridial (539 cells), and seven clusters resembling stem cells (10,686 cells) (Figure 1C).

Orthogonal body wall muscle fibres are transcriptionally distinct

Two muscle clusters were identified based on the expression of the well-described muscle-specific genes annotated as *Myosin*, *Troponin* and *Titin* (Smp_126240), as well as other differentially expressed markers; e.g. *collagen alpha 1(IV) chain* (Smp_159600)(Fig. 2A). High expression of *Myosin heavy chain* (*MYH*) (Smp_085540), an actin binding filament, was seen in both muscle cell clusters (Fig. 2A), and *in situ* hybridisation (Fig. 2B & C; Supplementary video 1) and image analysis (Supplementary video 2) highlighted 70 nuclei surrounded by *MYH* transcripts (i.e. ~19% of all cells in a miracidium are muscle cells). Counterstaining with phalloidin (which labels F-actin filaments) revealed an orthogonal grid of circular and longitudinal, but no diagonal, body wall muscles, and identified which *MYH*$^+$ nuclei belong to each muscle subtype. The nuclei of the circular muscles form two distinct bilaterally symmetrical lines that run peripherally from one pole of the larva to the other (~28 nuclei in total). Thirty-two more *MYH*$^+$ nuclei sit regularly spaced between the lines of the circular muscle nuclei and correspond to longitudinal muscles. At one pole of the larva,

another 12 *MYH*$^{+}$ nuclei form a unilateral cluster adjacent to the apical gland cell (identifiable by its four nuclei; Pan, 1980) (Fig. 2Cii; Supplementary video 1 and 2), but whether they belong to circular or longitudinal muscles is not clear.

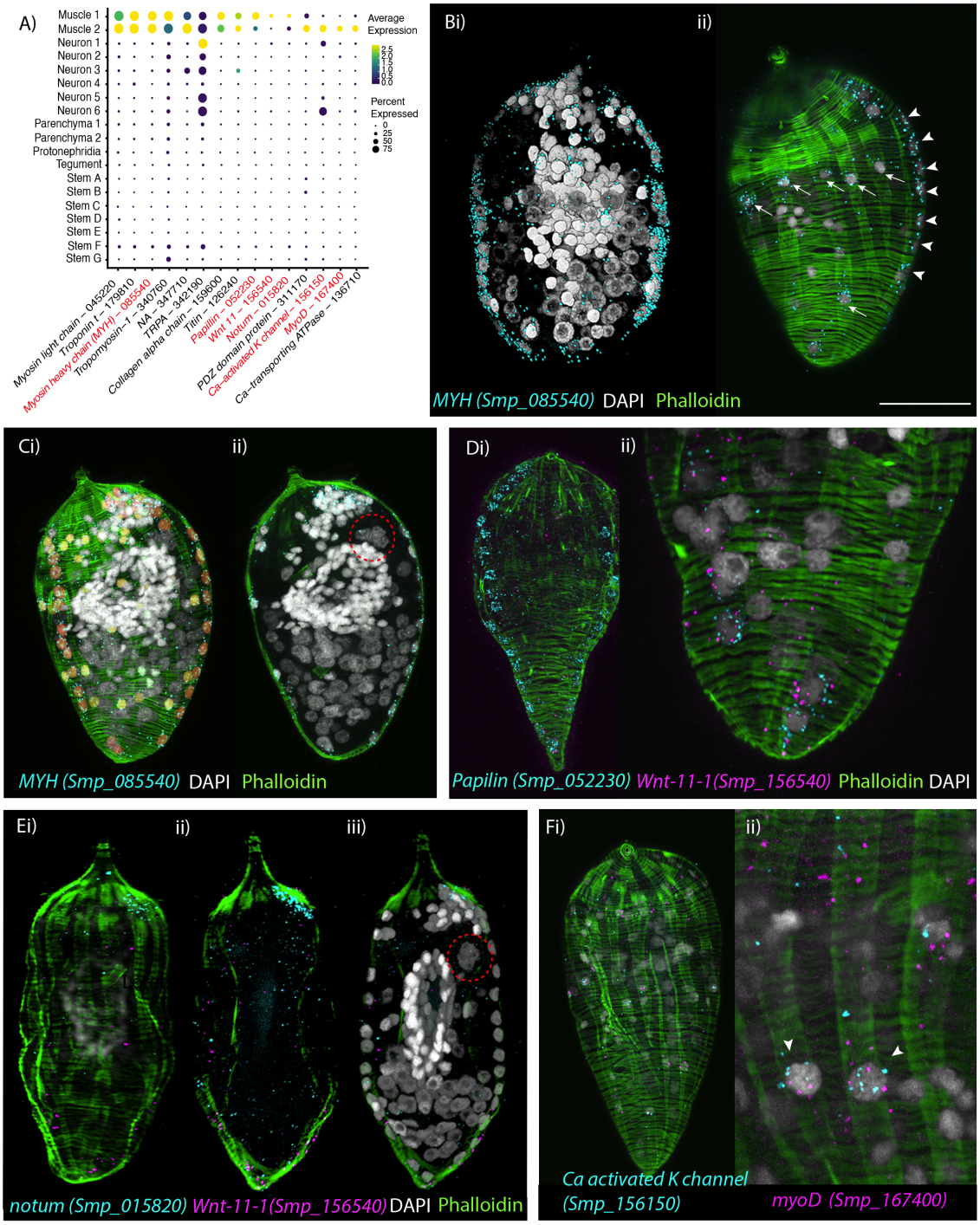


Figure 2. Orthogonal body wall muscles are transcriptionally distinct

A) Dot plot depicting the expression profiles of specific- or enriched- marker genes

in the two muscle clusters. Genes validated by *in situ* hybridisation (ISH) are marked in red. **B) i-ii)** Wholemount ISH (WISH) of myosin heavy chain *MYH*$^+$ and counterstaining with phalloidin reveals the nuclei of the circular body wall muscles (BWM) form two distinct bilaterally symmetrical lines that run peripherally from the pole of the larva to the other (arrowheads in ii). The longitudinal BWM nuclei (ii - arrows) are spaced regularly between the lines of the circular BWM nuclei and their actin fibres run orthogonal to the circular muscles. Scale = 20 µm. **C) i)** Segmentation of the *MYH*$^+$ cells shows 70 muscle cells in total; ~28 circular (segmented in orange), 32 longitudinal (yellow) and **ii)** 12 in an anterior unilateral cluster adjacent to the apical gland cell (identifiable by its four nuclei, inside dashed red circle). **D)** Expression of markers for muscle 1 cluster; **i)** *papilin* is expressed in circular BWM and *wnt-11-1* is expressed in 7 *papilin*$^+$ cells at the posterior pole; **ii)** close-up of posterior end of another miracidium showing the co-expression of *papilin* and *wnt-11-1*). **E)** i-iii) *notum* - an inhibitor of *wnt* signalling, is expressed highly at the opposite pole to *wnt-11-1* in the cluster of 12 muscle nuclei adjacent to the apical gland cell (nuclei in dashed red circle). **F)** Expression of markers for muscle 2 cluster; a *calcium activated potassium channel* (Smp_156150), and the transcription factor *myoD* (Smp_167400), show co-expression in longitudinal BWMs (arrowheads).

Muscle cluster 1 (2,407 cells) was distinguished from muscle cluster 2 by markedly higher expression of *papilin* (Smp_052230)(Fig. 2A), and ISH showed *papilin* transcripts along the two peripheral lines of circular muscle nuclei (Fig. 2D). Subpopulations of muscle 1 also express orthologs of *Wnt-11-1* and *Notum* (an inhibitor of wnt signalling, Kakugawa *et al.,* 2015)(Fig. 2A). *Wnt-11-1* (Smp_156540) is expressed at one pole, in seven of the *papilin*$^+$circular muscle cells (Fig. 2D; Supplementary video 3). At the other pole, *notum* (Smp_015820) is highly expressed in the cluster of 12 muscle nuclei adjacent to the apical gland cell (Fig. 2E). *Notum* is also expressed, albeit weakly, in some nuclei of circular muscles, including those expressing *wnt-11-1* (Supplementary video 4). Posterior Wnt signalling and anterior Wnt inhibition is a commonly observed feature of anterior-posterior body patterning in animals (Peterson and Reddien 2009). The expression of *wnt-11-1* at one pole suggests this is the posterior end of the larva, with *Notum* expression at the anterior. Furthermore, the unilateral, asymmetrical expression of *Notum* at the anterior end is suggestive of a dorso-ventral axis and/or a left-right axis. Neither axis has been identified in the miracidium.

Muscle cluster 2 (1,355 cells) was distinguished by higher expression of a genes annotated as *PDZ and LIM domain protein 7* (Smp_311170) and *flagellar attachment zone protein 1* (Smp_304000), as well as unique expression of a *calcium activated potassium channel* (Smp_156150), and the transcription factor *myoD* (Smp_167400)(Fig. 2A); ISH of the latter

two genes shows expression only in the longitudinal body wall muscles (Fig. 2F; Supplementary video 5). Orthologs of other axial patterning genes, *netrin* (Smp_172470), *wnt5* (Smp_145140), and a secreted frizzled *sFRP-2* (Smp_062560)(Witchley *et al*., 2013), show distinct expression in muscle 2 cells. Together, muscle 1 markers are more highly expressed in circular BWM, and muscle 2 genes are expressed in the longitudinal BWM, highlighting differences in the transcriptomes of these two types of myocyte that make up the orthogonal body wall musculature. Differential expression analysis shows that *wnt-11-1* (circular) and *myoD* (longitudinal) are among the most differentially expressed between the two muscle clusters (Supplementary table 4).

Neural abundance and diversity

We uncovered five clusters expressing the neural markers *complexin* (*cpx*; Smp_050220), *prohormone convertase 2* (*PC2*, Smp_077980) and Smp_068500 (a neural marker in adult worms, Wendt *et al*., 2020), and one that expresses *p25 alpha* (Smp_097490) a marker for ciliated neurons in adults (Wendt *et al*., 2020)(Fig. 3A). We numbered the neuron clusters 1-6. We identified distinct molecular fingerprints for five of the six populations at this resolution (Fig. 3A). *Complexin* was expressed around >165 nuclei, indicating that at least 45% of cells in the miracidium are neurons. 90 cpx^{+} nuclei form the nuclear rind of the brain (or neural mass/ring), and the remaining 75 are situated peripherally, either anterior or posterior to the brain (Fig. 3B). Sensory organelles were identified for all peripheral neural cell types.

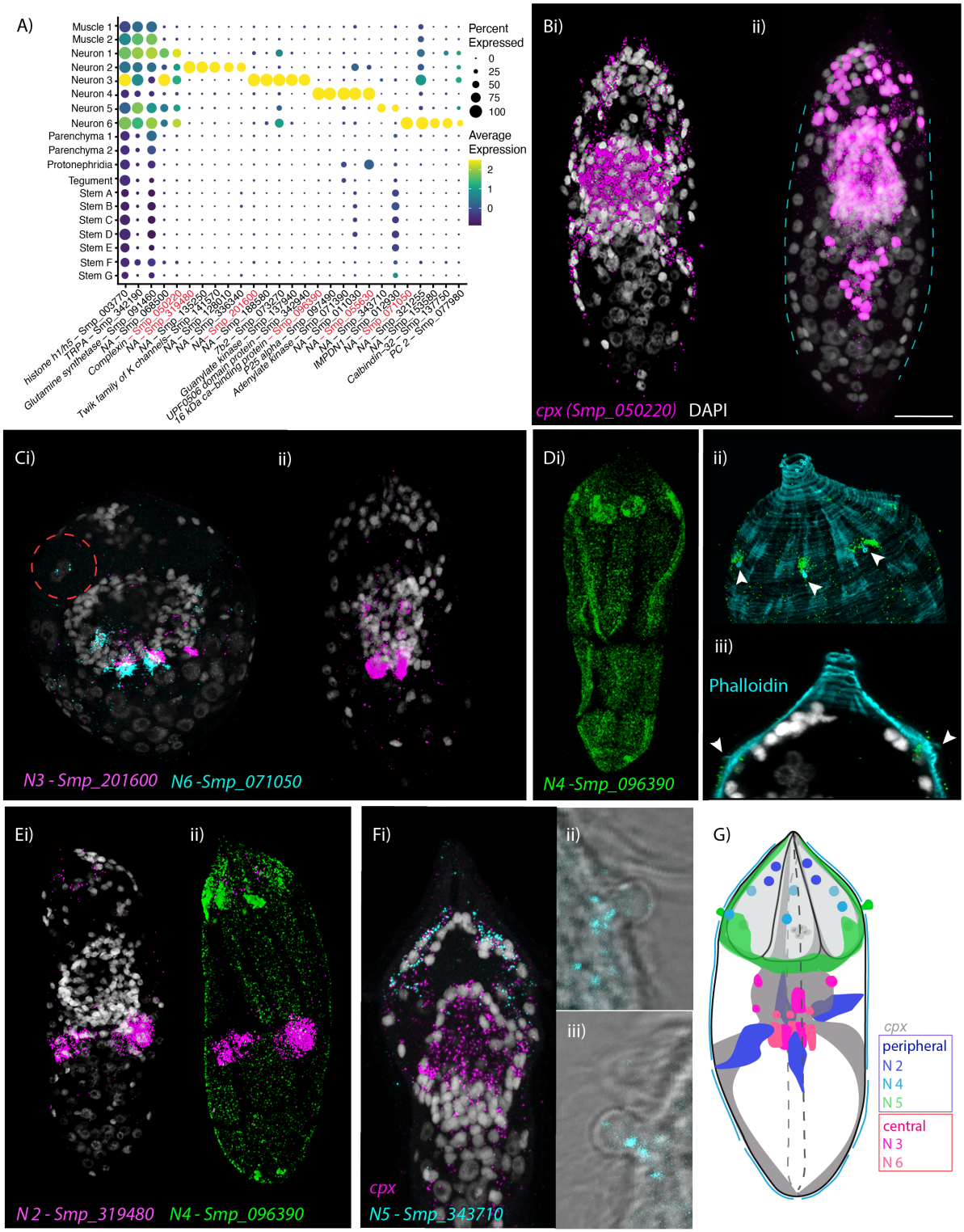


Figure 3. Neural complexity in a simple larva

A) Dot plot depicting the expression profiles of specific- or enriched- marker genes in the six neuronal clusters. Genes validated by *in situ* hybridisation (ISH) are marked in red. *Complexin* is expressed in all six clusters and is a neural marker in the adult (Wendt et al., 2020). **B)** i & ii) WISH and segmentation of *cpx*$^{+}$ cells reveals that at least 165 of the 365 cells in the miracidium are neural (Maximum intensity projection (MIP) **(i)** dorso-ventral

view and **(ii)** lateral view with nuclei of *cpx*$^{+}$ cells segmented in magenta, blue dashed line runs parallel to the circular muscle nuclei). The majority form the nuclear rind of the brain with the remaining either anterior with projections around the three gland cells or in two posterior clusters that send projections to the posterior pole perpendicular to the circular muscle cell nuclei. **C)** MIP of multiplexed ISH for top markers for neuron 3 (*Smp_201600*) and neuron 6 *(Smp_071050)* which are expressed in the multiple cells in the brain, including cells adjacent to each other. *Smp_201600* is also expressed in the apical gland cell (red dashed circle). **D) (i)** A top marker for neuron 4, a calcium binding protein *Smp_096390*, shows transcripts expressed in all the ciliary plates and in 6 cells towards the anterior pole (MIP). **(ii & iii)** Counterstaining with phalloidin shows that the nuclei of these anterior cells sit beneath the body wall muscle and send a projection externally between the 1st and 2nd tier of ciliary plates (arrowheads). **E) (i)** The top marker for neuron 2 *Smp_319480* is expressed in 4 cells anteriorly and ~22 that sit lateral and just posterior to the brain, these later cells send projections into the brain and out to the body wall at the latitude between the 2nd and 3rd tier of ciliary plates **(ii)** seen here expressing the neuron 4 marker gene *Smp_096390* (arrowheads). **F)** Neuron 5 marker *Smp_343710,* a putative calcium sensor, is expressed in **(i)** X cells whose nuclei sit outside of, and anterior to, the brain, in and around the paired lateral glands and their secretory ducts, and **(ii & iii)** in a pair of bilaterally symmetrical bulbous protrusions. **G)** Summary schematic of the *in situ* expression of marker genes for the neural cell clusters.

Cells from neuron clusters 1, 3, and 6 have nuclei in the nuclear rind of the brain. Neuron cluster 1 is large (1,470 cells), connects to stem F (below) and expresses the neuron-specific genes *complexin*, *7b2* and Smp_068500. At this resolution there were no clear marker genes restricted to neuron 1; however, using the self-assembling manifold (SAM) algorithm (Tarashansky *et al*., 2019) multiple subclusters could be resolved (Supplementary figure 1; Supplementary table 5). Three of the subclusters have a neuropeptide precursor (NPP) gene in the top 5 markers – *PWamide* (Smp_340860) (Wang *et al*., 2016)*, neuropeptide F* (Smp_088360)*,* and *neuropeptide Y* (Smp_159950) – and each of these genes shows expression in the nuclear rind of the brain (Supplementary figure 2). Neuron 3 is a distinct cluster composed of 282 cells that show high expression of the uncharacterised gene Smp_201600. ISH revealed three strongly Smp_201600^{+} cells, with nuclei sat in a median posterior location of the brain, and three cells with weaker expression, in a more anterior position (Fig 3C, supplementary video 6). Other top markers include the neuroendocrine protein *7B2* (Smp_073270)*, guanylate kinase* (Smp_137940) and a putative *orthopedia homeobox* protein (Smp_126560). Neuron 6 is a small cluster of 47 cells, expressing *complexin* and *7B2*. The uncharacterised gene, Smp_071050 is unique to, and highly expressed throughout this population (Fig. 3A). ISH showed expression in ~6 cells with

projections that extend into the brain and to the periphery, three of which sit adjacent to three neuron 3 cells. There is also weak expression in the four nuclei of the apical gland cell (Fig. 3Ci, Supplementary video 6). STRING analysis of neuron 6 markers produced two networks with functional enrichment, these were associated with 'synapses' and 'calcium'. In neuron clusters 1, 3 and 6, there was an enrichment for genes annotated with Gene Ontology (GO) terms related to 'neuropeptide signalling pathway' and 'neurotransmitter transport' (Supplementary table 6). There are, therefore, at least 3–6 transcriptionally distinct cell types in the brain mass.

Nuclei that are peripheral to the brain were observed for neuron clusters 2, 4 and 5. Neuron 4, found almost exclusively in the unsorted samples. It was composed of 40 cells with little or no expression of the pan-neural marker *cpx*, but each expressing the top marker Smp_096390, a 16 kDa calcium binding protein (*CaBP)*. *In situ* hybridisation revealed that *CaBP* is expressed in two very different types of cell: the ciliary plates (anucleate epithelial cells with thousands of motile cilia used for swimming) and six previously undescribed cells that we hereafter term sub-muscular cells. The latter cells send a projection out to the exterior between the first and second tier of ciliary plates but have nuclei sat below the body wall musculature (Fig. 3D)(Supplementary video 7 & 8). Another top marker, *p25 alpha* (Smp_097490*)* (a tubulin polymerization promoting protein associated with the basic function of cilia, Orosz & Ovádi, 2008), was also expressed in the ciliary plates and submuscular cells (Supplementary figure 2C). Given the association of *p25 alpha* with cilia in adults (Wendt *et al*., 2020), expression in the sub-muscular cells suggests they project cilia into the external environment and may be sensory cells. We investigated further to determine whether neuron 4 cells are ciliary plates, submuscular cells or both. ISH revealed that the pan-neural marker *complexin* is only co-expressed with *CaBP* in the submuscular cells, not the ciliary plates (Supplementary figure 2D), and *in silico* analysis revealed only one cell in the cluster that co-expresses *cpx* and *CaBP* (Supplementary figure 2E). Thus, neuron 4 is likely to be composed of ciliary plates. Consistent with this observation, neuron 4 markers were associated with GO terms related to 'mitochondrial transmembrane transport' and 'oxidoreductase activity' (Supplementary table 6) and STRING analysis showed two interaction networks: one associated with 'carbon metabolism and mitochondria envelope', the second associated with 'cytoplasmic dynein complex' and 'dynein light chain superfamily' (Supplementary figure 3). These predicted networks and their enriched

annotations are indicative of ciliated cells with high energy consumption, such as cells with thousands of motile cilia that beat in an ATP-dependent manner (Sanchez *et al.*, 2011).

Neuron 2 formed a large (1,140 cells) discrete cluster. The top marker was an uncharacterised gene, Smp_319480 (Fig. 3A) that, based on protein-structure prediction using I-TASSER (Yang & Zhang 2015), is similar to human mTORC1 (PDB entry 5H64; TM-score 0.85). *In situ* expression revealed ~26 *Smp_319480*$^{+}$ cells; four at the anterior end that have projections leading to the body wall in the first tier of ciliary plates, and ~22 that sit lateral and directly posterior to the brain. These latter cells formed four clusters of ~3–7 cells, with *Smp_319480* transcripts distributed around the nuclei and along projections leading into the brain and extending out to the body wall terminating between the second and third tier of ciliary plates (Fig. 3E. Supplementary video 9 and 10). Twenty-three multiciliated nerve endings form a girdle around the larva between the second and third tier of ciliary plates (Supplementary figure 2G) and Smp_319480 transcripts are expressed at the base of these cilia (Supplementary figure 2H); some transcripts also reach the nuclei of circular body wall muscles (Supplementary figure 2I), suggesting these cells may function as sensory-motor neurons. Other top markers include those encoding a putative TWIK-family of potassium channel (Smp_141570), two proteins with Polycystic Kidney Disease domains *(Polycystin-2-like Smp_165660* and *Polycystin-1-related Smp_165230*), a protein kinase A (Smp_079010), a GTP-binding protein (Smp_045990) and fibrocystin (Smp_303980). The GO terms 'regulation of apoptotic process' and 'potassium channel activity' were significantly enriched and STRING analysis showed networks enriched in 'plasma membrane bound cell projection'. Neuron 5 cluster comprises 276 cells. An uncharacterised gene, Smp_343710, is highly expressed in 55% of these cells. Based on protein-structure prediction, Smp_343710 is similar to a Ca2+ Sensor (PDB entry 3O78, TM-score 0.79, I-TASSER). ISH shows expression in 10-20 cells whose nuclei sit outside of, and anterior to, the brain, in and around the paired lateral glands and their secretory ducts (Supplementary video 11). There is also expression in a pair of bilaterally symmetrical bulbous protrusions (called lateral sensory papilla and thought to be depth sensors, Pan, 1980) (Fig. 3F). The GO terms 'calcium ion binding' and 'second-messenger-mediated signalling' were significantly enriched and STRING analysis revealed networks associated with 'synapses', 'ion channel binding' and 'cation transmembrane transporter activity' (Supplementary figure 3).

Protonephridia

There were 536 cells in this cluster, expressing genes common to excretory systems in other animals (Gąsiorowski *et al.,* 2021), e.g., the structural protein *nephrin* (Smp_130070) and the transcription factor *odd-skipped-related* (Smp_152210). 80% of the cells expressed the protonephridia marker *sialidase* (Smp_335600) (Sarfati *et al.*, 2021) (Fig. 4A). *Sialidase* was expressed around six nuclei in a bilaterally symmetrical pattern (Fig. 4Bi) (Supplementary video 12). On each side, sialidase transcripts were distributed along the s-shaped path of an excretory tubule, passing its nucleus, and connecting the anterior and posterior flame cells (Fig. 4Ci). Within the flames cells, expression was clearly seen around the nuclei, rather than the barrel (Fig. 4Cii). The ciliated flame cells and excretory ducts comprise the worm's excretory system. Accordingly, STRING (Supplementary figure 4) and GO analyses (Supplementary table 6) of this cluster confirmed enrichment for 'cilium organisation*'* and 'cytoplasmic dynein complex*'*, and 'cytoskeleton organization*'* and 'microtubule-based process*'*, respectively.

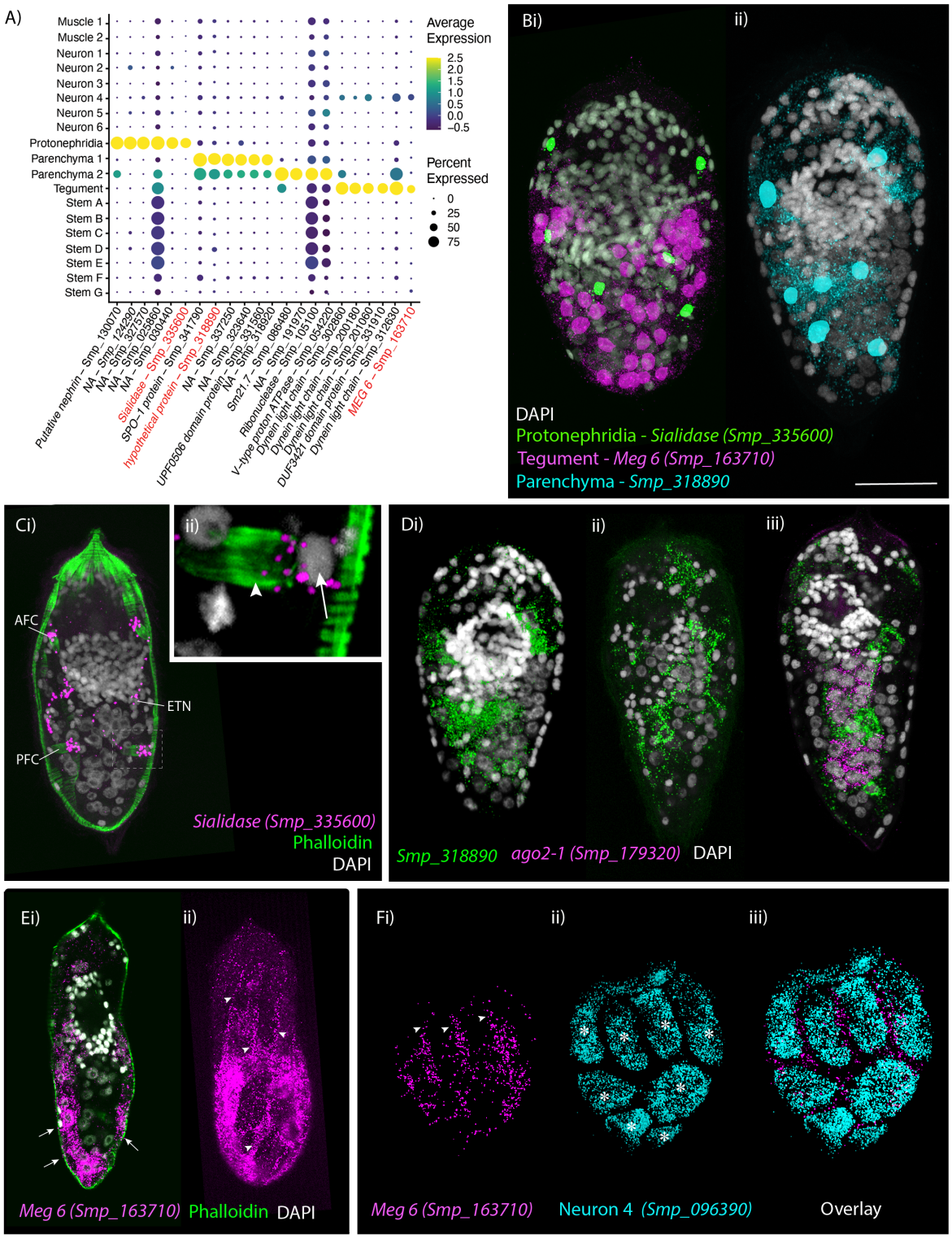


Figure 4. Identification of the protonephridia, parenchymal and tegumental cells in the miracidium. **A)** Dot plot depicting the expression profiles of specific- or enriched- marker genes in these cell type clusters. Genes validated by *in situ* hybridisation are in red. **Bi & ii)** WISH and segmentation of marker genes show that there are 6 cells that express *sialidase* (a marker for protonephridia), 46 cells that express *Meg-6* (a marker for tegument) and 7 cells that express an uncharacterised gene, *Smp_318890* (a marker for parenchymal cells). **Ci)** The

protonephridial marker *sialidase* shows s-shaped expression with transcripts extending from the nucleus of the anterior flame cell (AFC) along the excretory tubule and it's nucleus (ETN) to the posterior flame cell (PFC), **Cii)** where it is expressed around the nuclei (arrow) rather than the barrel (arrowhead) of the flame cells. **Di)** The pan-parenchymal marker, *Smp_318890*, was expressed in 2 anterior cells (one either side of the brain) and 5-7 cells posterior to the brain. **Dii)** These cells have long cytoplasmic protrusions that reach between all the other cells, **Diii)** including the *ago2-1*$^{+}$ stem cells. **E & F)** The tegument marker *Meg-6* shows expression around nuclei in the posterior two-thirds of the larvae. The nuclei are below the body wall muscle, and cytoplasmic protrusions reach between muscle filaments (arrows) and form the epidermal ridges (arrowheads) between the ciliary plates (which are visible in **Fii & iii** expressing the neuron 4 marker *Smp_096390*) (asterisks).

Parenchyma

We identified two clusters of cells (1,093 cells, parenchyma 1; and 605 cells, parenchyma 2) that are transcriptionally similar to the parenchymal cells found in later developmental stages; e.g. they express *villidin* (Smp_063330), *cathepsin B-like cysteine proteinase* (Smp_141610) and *serpin* (Smp_090080) (Diaz *et al.,* 2020; Wendt *et al.*, 2020). All top marker genes for parenchyma 1 were also expressed in parenchyma 2 but at lower levels, whereas the uncharacterised gene, Smp_191970*,* was uniquely expressed in parenchyma 2 (Fig. 4A). A further uncharacterised gene, Smp_318890, was a top marker for both clusters, with expression confirmed *in situ* within 7–9 cells; two of which are located anteriorly on either side of the brain and 5-7 located posterior to the brain (Fig. 4B & D). These parenchyma cells have long cytoplasmic protrusions that reach between all the other cells, filling the intercellular space (Fig. 4Dii-iii; Supplementary video 13 & 14). We performed ISH for two parenchyma 2 markers, Smp_191970 and Smp_320330, but the first showed no detectable signal and the second had broad expression across all cells with no specific overlap with the pan-parenchymal marker Smp_318890*.* Therefore, by ISH, we were only able to identify parenchymal cells in general and not cells belonging to the two sub-clusters. STRING analysis for the parenchyma 1 showed enrichment in networks for 'carbohydrate metabolic process' and 'hydrolase', and parenchyma 2 'phagosome', 'mTOR signalling pathway', 'mitochondrial inner membrane' and 'proton-transporting two sector atpase complex' (Supplementary figure 5). Analysis of GO annotations was able to distinguish between the two parenchyma sub-populations, with parenchyma 1 genes enriched for 'negative regulation of endopeptidase activity', 'sphingolipid metabolic process' and 'sodium ion transport', and parenchyma 2 genes enriched for processes that are 'glycolytic' and 'microtubule-based' as well as 'proton transmembrane transport'.

Tegument

We identified a cluster of 366 cells as tegumental, based on expression of markers for tegument in the adult worm (Wendt *et al.,* 2020); e.g. genes encoding *dynein light chain* (Smp_302860, Smp_200180), *Natterin 3* (Smp_331910) and *MEG6* (Smp_163710). Two more dynein light chain genes were top markers (Smp_201060, Smp_312630), and a *tetraspanin* (Smp_303290) and *Gelsolin* (Smp_008660) were specific to this cluster. *Meg6* (Smp_163710) is expressed in 66% of cells in this cluster; ISH shows 46 *Meg6*$^+$ cells and the nuclei are located in the posterior two thirds of larva. The nuclei sit below the body wall muscle, and cytoplasmic protrusions reach between muscle filaments and form the epidermal ridges between the ciliary plates (Fig. 4E & F). A second tegument marker, a *dynein light chain* (Smp_200180), shows overlapping expression with *Meg6* (Supplementary video 15). Analysis of annotated GO terms was inconclusive, with enriched GO terms that included 'microtubule-based process', 'signal peptide processing', 'intramolecular oxidoreductase activity' and 'calcium ion binding'. However, STRING analysis revealed a clearer pattern with a network of 25 genes associated with protein localisation to the ER, folding, export and post translational protein targeting to the membrane (Supplementary figure 4).

There are two major stem cell populations, and they cluster further by sex

We identified stem cells based on the expression of known (Wang *et al*., 2013; 2018) stem cell markers for the mother sporocyst stage, e.g. *ago2-1* (Smp_179320), *nanos-2* (Smp_051920), *fgfrA* (Smp_175590), *fgfrB* (Smp_157300), *klf* (Smp_172480) (Fig. 5A). The sporocyst has three stem cell populations: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*fgfrA*$^+$, *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$, *fgfrB*$^+$ and *hesl*$^+$)(Wang *et al*., 2018; Safarti *et al*., 2021). In the miracidium, we identified seven stem cell clusters by expression of *ago2-1*: stem A and B that are *klf*$^+$ and *nanos-2*$^+$ (i.e. Kappa-like); stem C and D that are *nanos-2*$^+$, *fgfrA*$^+$, *fgfrB*$^+$, and *hesl*$^+$ (i.e. they resembled both Delta and Phi)(Fig. 5A, Supplementary figure 6), stem E that forms a 'bridge' between the stem A, B, C and D; stem F that connects the stem clusters to neuron 1 and tegument; and a small stem G cluster that is an offshoot from stem F (Fig. 5B).

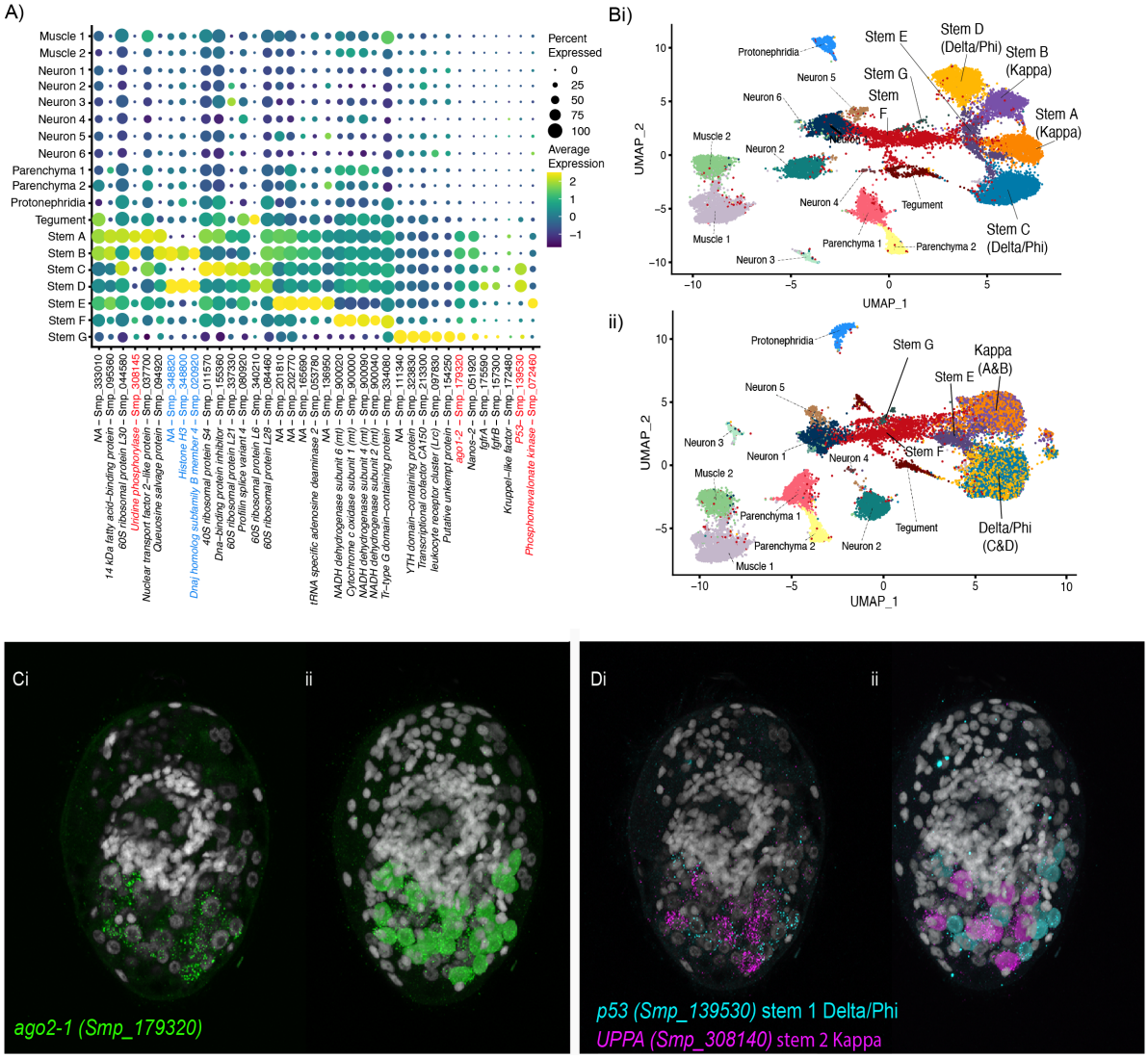


Figure 5. Two defined populations of stem cells that cluster further by sex.
A) Dot plot depicting the expression profiles of specific- or enriched- marker genes in the stem cell clusters. Genes specific to the W (female) sex chromosome genes are highlighted in blue. Genes validated by ISH are marked in red. **Bi)** UMAP including all genes shows that there are two Delta/Phi and two Kappa stem clusters, **ii)** UMAP showing that removal of all genes specific to the W and Z sex chromosomes results in one Delta/Phi and one Kappa cluster, indicating that the stem cells are transcriptionally different in male and female miracidia due to the expression of sex-linked genes. **C&D)** Multiplexed ISH showing three stem cell markers simultaneously in the same individual: *ago2-1* (Smp_179320) (pan-stem), *p53* (*Smp_139530*) (Delta/Phi) and *Uridine phosphorylase A* (*UPPA*, Smp_308140) (Kappa). **Ci)** *Ago2-1* expression reveals stem cells lateral and posterior to the brain (Optical section), **ii)** MIP with segmentation of the 23 *ago2-1*$^{+}$ cells. **Di)** *P53* and *UPPA* expression shows the two stem cell populations intermingled. **ii)** MIP and segmentation reveals there are more *p53*$^{+}$ cells than *UPPA*$^{+}$ cells (in this larva 15 are p53^{+} and 9 are UPPA^{+}).

The Delta/Phi-like and Kappa-like clusters are each composed of two well-separated clusters that can be distinguished based on differential expression of discriminatory markers located on the W (female-only) sex chromosome (Fig. 5A). The sex chromosomes are composed of sex-specific regions that are unique to each chromosome flanked by pseudoautosomal regions (PARs) that are common to both Z and W. The W-specific region (WSR) is a large highly repetitive region that contains 38 genes, most of which are gametologues (genes with homologues in the opposite sex-specific region). The Z-specific region (ZSR) contains 941 genes, but only 35 of which are gametologues (Buddenborg *et al.*, 2021). In our mixed-sex dataset, cells that express WSR genes (W^+ cells) should come from female parasites, and cells with no detectable expression of WSR genes (W^- cells) should come from males; however, the possibility of female cells with signal dropout (Kharchenko *et al.,* 2014) for the WSR genes cannot be excluded. This suggests that the Delta/Phi-like and Kappa-like clusters are split by sex (Fig. 5A and Supplementary figure 7). Removing from the dataset genes located on the WSR, ZSR, or just the gametologues, and reclustering did not impact the overall UMAP clustering, i.e. stem cell clusters remained split by sex (Supplementary figure 7). However, when all 929 ZSR and WSR-genes detected in these data were removed simultaneously from the cluster analysis, the two Delta/Phi-like and two Kappa-like clusters each collapsed into single stem cell clusters, giving 5 stem cell clusters in total (i.e. Delta/Phi, Kappa, E, F and G; Fig. 5B). This indicates that the expression of sex chromosome genes are drivers of the clustering.

In many animals, a process called dosage compensation equalises gene expression across sex chromosomes that are unevenly distributed between the sexes. The proportion of reads mapping to either the WSR or ZSR genes across all clusters shows that the male stem cells (i.e. those that lack the WSR but have two copies of the ZSR) have a relatively higher expression of the ZSR genes compared to those in female cells (Supplementary figure 8). This suggests an incomplete or a lack of dosage compensation in the male Delta/Phi and Kappa stem cells. The combinatorial effect of high specific expression of WSR genes in female Delta/Phi and Kappa stem cells and an incomplete dosage compensation of ZSR genes in the Delta/Phi and Kappa stem cells may be the main drivers for these stem cells clustering by sex. Each of the two phenomena is sufficient and enough to split these stem clusters in two. The splitting of clusters based on the effect of WSR and ZSR gene expression is limited to Delta/Phi-like and Kappa-like cells, and not found in somatic and other stem clusters. However, when the cells were pseudo-typed based on expression of WSR genes (i.e.

only found in female animals), the percentage of reads mapping to the ZSR was higher in the W^- typed cells across the majority of the cell clusters (Supplementary figure 8).

To locate the stem cells *in situ*, we investigated the expression of *ago2-1*. ISH revealed *ago2-1*$^+$ cells distributed lateral and posterior to the brain with large nuclei and prominent nucleoli (Fig. 5Ci). There is variation in the number of *ago2-1*$^+$ cells between larvae, ranging from 20-29 (n = 5, median = 23, mean = 23.4)(i.e. ~7% of cells). Cells in the Delta/Phi-like cluster express a canonical *p53 tumour suppressor family gene* (Smp_139530) as a top marker; *in situ* expression shows 12–14 *p53*$^+$ cells (n=3)(Fig. 5D). Cells in the Kappa-like cluster express *Uridine phosphorylase A* (*UPPA*, Smp_308145) as a top marker; *in situ* expression shows 8–9 *UPPA*$^+$ cells, and they are located amongst the *p53*$^+$ cells (n = 3)(Fig. 5D). We were not able to detect cells from the stem E, F and G clusters in the miracidium using ISH. *Phosphomevalonate kinase* (*pmk*) (Smp_072460) is a predicted marker for the stem E cluster, and is highly expressed in 49% of the cells; however, it is also expressed at lower levels in subpopulations of Kappa and Delta/Phi (Fig. 5A). Even when multiplexed as *p53/UPPA/pmk*, no *pmk*$^+$-only cells could be found (Supplementary video 16). Likewise, there are no predicted marker genes that are unique to the stem F and G clusters and in multiplexing experiments, using the pan-stem marker *ago2-1* with markers for Delta/Phi *(p53)* and Kappa *(UPPA)*, no *ago2-1*$^+$-only cells could be found (Supplementary video 17). Therefore, we were only able to spatially confirm two stem cell populations within the miracidia: Delta/Phi and Kappa. Cells in stem E, F and G *in silico* clusters may be in transcriptionally transitional states or stressed/damaged/dying cells. Stem F cells are largely from the unsorted samples, whereas the majority of Stem G cells are from the sorted samples.

EdU labelling of miracidia shows that there is no incorporation into cells during the free-swimming stage (Supplementary figure 9), indicating that no DNA duplication takes place in the stem, or other, cells during this time. As the first stem cells start proliferating in the mother sporocyst ~20 hours post-infection (Wang *et al.*, 2013), this suggests that both stem cell populations pause cell division for at least 26 hours.

Many top markers for both stem clusters are ribosomal proteins e.g. *Smp_011570* and *Smp_044580* (Fig. 5A), and STRING analysis shows that most of the top 100 markers of each cluster both form large single networks enriched for genes involved in translation,

ribosome biogenesis and gene expression (Supplementary figure 10). This is particularly striking in the Delta/Phi cluster, where 46 of the top 100 markers are associated with ribosome biogenesis and translation. Amongst the top 20 genes, differentially expressed between Kappa and Delta/Phi (Supplementary figure 11A; Supplementary table 7), were six transcription factors up-regulated in Delta/Phi: orthologs of *homothorax* (Smp_147790), *lin-39* (Smp_247590), *tiptop* (Smp_180690) and *hes-related* (Smp_024860), as well as two that are critical in tegument generation in the adult worms, a canonic*al p53* (Smp_139530) and a parasitic flatworm-specific transcription factor *Zfp1* (Smp_145470) (Wendt *et al*., 2018; 2022). Enriched, or specific to, Kappa cells were genes involved in glycolysis and lipid metabolism i.e. malate dehydrogenase (Smp_129610), 14 kDa fatty acid-binding protein (Smp_095360) and elongation of very long chain fatty acids protein (Smp_051810). GO analysis of differentially expressed genes (with adjusted p-values <0.001; Supplementary table 8) revealed that the most significantly upregulated genes in Delta/Phi (573 genes) were related to structural component of the ribosome and translation, while up-regulated genes in Kappa (948 genes) are involved in transcription and metabolism (Supplementary figure 11B). Overall, Delta/Phi cells were characterized with up-regulation of protein translation capacity and the simultaneous co-expression of multiple transcription factors, in comparison with Kappa stem cells which have higher levels of glycolytic & fatty acid metabolism gene expression.

Discussion

We have shown that the miracidium is composed of only ~365 cells, and we used scRNA-seq of 20,478 cells to identify 19 transcriptionally distinct cell types. Multiplexed *in situ* validation of cell type marker genes enabled the cell types to be mapped back onto the larva and has revealed where these cells sit in relation to each other. Furthermore, the nuclear segmentation of the cells expressing tissue-specific genes has revealed the relative contribution of each tissue type to the animal.

Nineteen cell types may be a conservative characterisation, but at this resolution most cell types formed clearly defined clusters (the exceptions being neuron 1 and stem E, F and G). There was sufficient sensitivity to detect cell types with as few as six cells per larva e.g. neuron 3 and 6, and the protonephridial system. We predict some rarer cell types might be embedded in existing clusters, particularly in neuron 1, and a complete neural cell atlas and

connectome of the 165 neural cells is a follow up study in progress (Rawlinson *et al*. in prep). Some cell types were preferentially recovered. In particular, ISH revealed stem cells comprise only 7% of the total cell number in the larva, yet they accounted for ~52% of the scRNA-seq data. Preferential recovery may be due to differences in the robustness of different cell types to dissociation, cell sorting, and/or the microfluidics of the 10X GEM capture process. For instance, the neuron 4 (the ciliary plates) cell type was only found in sufficient numbers in unsorted samples to be defined as a cluster. Therefore, there are clear benefits to including unsorted cell samples to discover the diversity of cell types in a tissue, organ, or organism, especially when cell types may be fragile, sticky, or unusually-shaped. Even with the inclusion of unsorted cells, some cell types were not obviously captured: the two lateral, and the apical, gland cells. These cells were easily identified after dissociation by their distinct flask-shape, but were not identifiable in these data. Perhaps, because of their large size (20x50 µm) and irregular shape they were not incorporated into the 10X GEMs, which were primarily optimised for mouse and human tissues.

The *S. mansoni* miracidium, with its diversity of differentiated cells present in relatively low numbers, is a simple larva and makes a good system to investigate the biology and interactions of cells and their transcriptomes across the life cycle and between species. Compared to the ~365 cells of the miracidium, there are ~1,600 cells in the cercariae and ~900 in the schistosomula (Diaz *et al.*, 2020); in other species, the L1 larvae of *C. elegans* has 550 non-gonadal nuclei at hatching (Sulston and Horvitz, 1977), and a three-day-old *Platynereis* larva has 8,852 cells (at 72 hours post fertilisation)(Veraszto *et al*., 2020). The two muscle clusters in our dataset correspond to the circular and longitudinal muscles that comprise the orthogonal body wall muscles of the miracidium. Later life cycle stages have 3-8 muscle clusters (Diaz *et al.*, 2020; Wendt *et al*., 2020), which may correspond to the greater diversity of muscle subtypes that develop over ontogeny; e.g. diagonal body wall muscles, missing in the miracidium, are found in the cercariae and intra-mammalian developmental stages (Collins *et al*., 2011; Mair *et al*., 2000). We identified differences between circular and longitudinal muscle transcriptomes that have also been observed in planaria; e.g. *MyoD* and *wnt-11-1* are expressed in longitudinal and circular muscles, respectively (Scimone *et al*., 2017). The axial patterning genes expressed in planaria muscles are essential for patterning during regeneration (Witchley *et al*., 2013; Vogg *et al*., 2014). The expression of orthologs in *S. mansoni* muscles, a nonregenerative animal, suggests that these genes may regulate schistosome stem cells during homeostasis in the adult (Wendt *et al*., 2020); however, in the

miracidium they may provide positional information for axial patterning of the daughter sporocysts developing within the mother sporocysts.

The nervous system accounts for nearly 50% of cells that make up the larva and is the most transcriptionally heterogeneous tissue, as has been observed in other animals (Siebert *et al*., 2019; Vergara *et al*., 2021). At the end of embryogenesis, at hatching, most of the miracidia neurons are fully differentiated and form the functional neural circuits of the larva, which is evident in our scRNAseq data by the presence of discrete, well-defined neural cell clusters. The one exception being neuron 1 that connects to stem F and lacks distinct marker genes. This suggests that it may contain neural progenitors in an intermediate (transcriptionally transitional) state that are indistinguishable between stem cells and nerves at the microscopy level. Given that the miracidia for these experiments were collected four hours after inducing hatching, the larvae ranged from newly hatched to four hours post-hatching, and a range of differentiation states could therefore be encompassed within the samples. Although we have shown that there is no cell proliferation in the miracidium, the possibility remains that the transcriptional states of stem and neuron 1 cells differ between freshly-hatched miracidia and those collected four hours post-hatching.

In situ hybridisation of marker genes revealed many of the nerves described by Pan (1980) (e.g. bipolar ganglion cells and lateral sensory papilla), and examination of their transcriptional profiles indicates their function. For example, the location and shape of neuron 2 cells (revealed by ISH) resembles that of the bipolar ganglion cells with multi-ciliated nerve endings (Pan, 1980). Top markers for this cluster include a putative *mTORC1*, two Polycystic Kidney Disease (PKD) domain-containing proteins *(Polycystin-1-related* and *Polycystin-2-like),* a *GTP-binding ras-like* and *fibrocystin,* mammalian orthologs of which function in cells with primary cilia in the detection of flow sheer and regulate mTORC1 signalling to maintain the cell in a quiescent and functional state (Lai & Jiang, 2020). Polycystins PKD1-1 and PKD2-1 expressed in ciliated neurons of the annelid *Platynereis dumerilli* larva are involved in mechanosensation, detection of water vibrations and mediate the startle response (Bezares-Calderón *et al.,* 2018). This gene expression profile indicates that neuron 2 cells could function as flow and/or vibration sensors.

Many of our findings corroborate the putative functions of cell types ascribed using ultrastructural analysis by Pan (1980). For example, interstitial cells that contain abundant

glycogen particles and lipid droplets and have long cytoplasmic protrusions were proposed to function as sources of energy for the non-feeding larva and as nurse cells for the germinal/stem cells during the first few days post infection of the snail. For the parenchymal cells identified here, their long cytoplasmic protrusions reaching between all cell types, and transcriptomes enriched in genes involved in metabolism and energy generation, suggests they are the interstitial cells described by Pan (1980). Similarly, the transcriptome of the tegumental cells is consistent with the ultrastructural data, showing a predicted network of genes involved in protein folding, export and translocation. Named epithelial cells, Pan (1980) observed abundant RER, ribosomes, Golgi apparatus, and complex carbohydrates in these cells, suggesting they are active in production of exportable proteins. Once inside the snail the ciliary plates are shed and the tegumental cells form the new tegument of mother sporocysts. Pan (1980) also observed large numbers of membrane-bound vesicles that become part of the new tegument membrane that rapidly forms around the mother sporocyst during the early stages of transformation, speculating that the contents may protect the miracidia/mother sporocyst from attack by the host amebocytes.

Pan (1980) determined that there are about 20 stem cells in the miracidia. We show that there is variation between larvae, with the median being 23. We have demonstrated that these form two transcriptionally distinct populations, Delta/Phi and Kappa, which we have validated *in situ*. We show that both populations are non-dividing during this stage; however, their transcriptomes suggest they are in different activation states. The up-regulation of protein translation capacity, that is evident in Delta/Phi, is the first marker of a pre-activation stage in stem cells (Shin *et al*., 2015). Whereas the higher levels of glycolytic & fatty acid metabolism genes in Kappa are representative of quiescent stem cells rather than primed and activated stem cells (Llorens-Babadilla *et al*., 2015). This finding aligns with the discovery of two stem cell populations in the mother sporocyst that have different proliferation kinetics (Wang *et al*., 2013). The pre-activated state of Delta/Phi stem cells in the miracidia may signify that they are the first population to proliferate once the miracidium becomes the mother sporocyst inside the snail. The differential expression analysis between Delta/Phi and Kappa also suggests differences in cell potency between the populations. The simultaneous co-expression of multiple transcription factors in Delta/Phi (some of which are known to be lineage specific e.g. *p53* and *zpf1-1* in tegument development (Wendt *et al*., 2018; 2022), compared to Kappa, indicates a multipotent progenitor stage rather than pluripotency (Kumar *et al*., 2017). In contrast, the Kappa stem population may be pluripotent because it is thought

to generate the other stem cell types (Delta and Phi) in the daughter sporocyst stage, and then give rise to germline and posterior growth zone somatic stem cells in the intra-mammalian stage (Wang *et al*., 2018). Adding the miracidia stem cells to this scenario suggests that a Kappa-like population can be traced from hatching to sexual maturity and includes the germline lineage, whereas the Delta/Phi lineage may provide supporting cell types to ensure the transmission of the Kappa stem line.

Sarfati *et al*., (2021) proposed that, unlike most animals that segregate their germline during embryonic development, the schistosome does not specify its germline until the intra-mammalian juvenile stage, by which point the parasite has already completed embryogenesis and gone through two additional postembryonic body plan remodelling processes (i.e. daughter sporocyst and cercariae development). Our data in the miracidia showing two distinct stem cell populations, one of which expresses *klf* (a homolog of KLF4, an important transcriptional regulator governing pluripotency; Takahashi & Yamanaka, 2006) suggest that cells that may go on to produce the germline are a distinct cell type at this stage and this putative pluripotent line is specified during embryogenesis.

The most unexpected finding regarding the stem cells was the further clustering of the two populations by sex. This is the first identification of sexual dimorphism in the miracidium. Remarkably, this degree of transcriptomic sexual dimorphism was most evident in the Delta/Phi and Kappa stem cell clusters, with no distinct clustering by sex observed in other cell clusters. This finding would suggest a tissue specific dosage compensation of ZSR-linked genes. However pseudotyping of the cells by sex showed generally higher levels of ZSR across most cell clusters; gene dosage compensation may, to some degree, also be incomplete in the somatic tissues of the miracidium. This is the first report of tissue specific dosage compensation in schistosomes, and follows investigations in gene dosage compensation across developmental stages using bulk RNAseq (Picard et al., 2019). A similar phenomenon has been described in mouse embryonic stem cells and in *Drosophila* where a precise regulation of gene-by-gene dosage compensation mechanism is critical to the proper development of specific tissues and organs, such as the wing and eye of the fly (Valsecchi et al.,2018). In contrast to the miracidium, sex differences in undifferentiated stem cells (neoblasts: somatic or germline stem cells) have not been reported in later stages of the schistosome life cycle (Wang *et al.,* 2018; Diaz *et al.*, 2020; Wendt *et al*., 2020). In fact, it may not have been possible to detect these differences in earlier studies, because scRNA-seq

data were mapped onto older versions of the genome, where the sex chromosomes were not as well defined. In addition to using a new improved reference sequence, the present study has also benefited from a high level of coverage, enabling differences in stem cells to be more easily discerned. Whether these sex-linked gene expression and dosage differences have functional implications in the development of male and female sporocysts remains to be investigated.

The current understanding of schistosome development is that differentiated somatic cells of the miracidium do not continue into the daughter sporocyst stage, the cercariae and beyond; it is only the stem cells that persist and differentiate into the cell types of these subsequent stages (Wang *et al*., 2018). In this respect, all miracidium somatic cells would be miracidium-only cells; however, our data show that their transcriptional fingerprints are similar to comparable cell types at later stages. The only truly 'miracidium-specific' cell type is the ciliary plate, which is not found in later life cycle stages. The cellular composition of an organism varies markedly over the course of development. Moreover, many cell types that develop during later stages are not present in the miracidia, including the esophageal and primordial gastrodermal cells that emerge during the development of the cercariae, plus germline, vitellaria and Mehlis gland cells that emerge during sexual maturation.

We have produced a cell atlas of the miracidium, the first free-swimming larval stage of the parasite *Schistosoma mansoni,* and we have recovered almost all cell types predicted by the existing body of literature. High coverage data from individual cells has allowed us to examine how gene expression changes across and within tissue types; we were also able to validate all these tissue types with ISH. Through these data we are able to show that even in this simple and sexually monomorphic short-lived stage, there are different populations of muscle, a diverse and distinct range of neural cell types, and that stem cell transcriptomes are already different based on the sex of the miracidium. The molecular definitions and spatial analyses of the miracidia cell types we have presented here now allow us to explore the molecular processes and pathways to reveal the cell functions and interactions. Already a picture is emerging of how the peripheral neurons may detect environment cues, connect to the brain, muscles and ciliary plates to orientate in the water column; how the parenchymal cells supply nutrients to the surrounding cells; and how the tegumental and stem cells are primed for the transition to the intramolluscan stage.

Methods

Parasite material

Schistosoma mansoni (NMRI strain) eggs were isolated from livers collected from experimentally-infected mice perfused 40 days post infection. All animal regulated procedures, including the mouse experimental infections at the Wellcome Sanger Institute were conducted under Home Office Project Licence No. P77E8A062 held by GR. All the protocols were presented and approved by the Animal Welfare and Ethical Review Body (AWERB) of the Wellcome Sanger Institute. The AWERB is constituted as required by the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

Nuclei count

Schistosoma mansoni eggs isolated from infected mouse livers were hatched under light in freshwater (Mann *et al.,* 2010), and the miracidia collected within four hours of hatching. The sample was put on ice for an hour to sink and concentrate the larvae and all but 1 mL of the water was removed. One mL of 8% paraformaldehyde (4% final PFA concentration) was added and the miracidia were fixed for 30 minutes at room temperature. The larvae were rinsed in PBS/Tween (0.1%) and mounted in DAPI Fluoromount for confocal imaging. The whole mounts were imaged on a Leica Sp8 using a 40X objective with the pinhole set to 1AU, and a 30-50% overlap between Z slices. The Z-stacks were imported into ImageJ and each nucleus was manually segmented and counted using the plugin TrakEM2 (Cardona et al 2012).

Single-cell tissue dissociation

Schistosoma mansoni miracidia for single-cell dissociation were collected as described above. Eggs isolated from mouse livers were hatched in nuclease free water (NFW) and collected within four hours of hatching. Two samples of miracidia were collected; sample 1 contained approximately ~8250 individuals, and sample 2 ~11,700. The samples (live miracidia in NFW in Low Protein Binding 15 mL Falcon Tubes [cat# 30122216]) were placed on ice for an hour to sink and concentrate the larvae and all but 20 µl of water was then removed.

Three hundred µl of digestion buffer (2.5 µl of 0.5 M EDTA in 300 µl of Hank's Buffered Salt Solution without Ca2+ and Mg2+ [HBSS-/-]) was added to the larvae. We reconstituted

250 mg of the enzyme Native *Bacillus licheniformis* Protease (NATE-0633, Creative Enzymes)(Adam *et al*., 2017) in 2.5 mL of HBSS-/- (to give a stock concentration 100 mg/ml) and added 200 µl of this stock to the larvae in buffer (final enzyme concentration = 40 mg/ml). The samples were kept on ice for 45 minutes and pipetted gently ~20 times every five minutes. Two hundred µl of 25% BSA in 1X HBSS-/- was added to the sample to quench the enzyme and 10 mls of ice-cold PBS +/+ was then added to bring the BSA to ~0.5%. The samples were centrifuged at 350 g for 5 minutes at 4°C, and all but 20 µl of the supernatant removed. We reconstituted 5 mg of the enzyme Liberase™ (Thermolysin Medium) Research Grade (LIBTM-RO Roche) in 2 mL of PBS (to give a stock concentration of 2.5 mg/ml) and added 100 µl of this stock to the sample (final enzyme concentration = 2.08 mg/ml). The samples were incubated at room temperature for 15 minutes, pipetting gently ~10 times every five minutes. The suspension was passed through 60 µm and 30 µm cell strainers (Falcon) and resuspended in 5 mL of cold 1× PBS supplemented with 0.5% BSA. The samples were centrifuged at 350 g for 10 minutes at 4°C, and all but 700 µl of the supernatant was removed.

Sample 1 cell suspension was sorted; 200 µl was used for a no-stain control for autofluorescence during cell counting and sorting. The remaining 500 µl of cell suspension was stained with 1 µg/ml of Propidium Iodide (PI; Sigma P4864) to label dead/dying cells. Sample 1 was sorted into an eppendorf tube using the BD Influx™ cell sorter (Becton Dickinson, NJ) by enriching for PI negative cells (i.e., live cells). It took ~3 hours from the beginning of the enzymatic digestion to generating single-cell suspensions ready for library preparation on the 10X Genomics Chromium platform. Sample 2 cell suspension was not sorted.

<u>10X Genomics library preparation and sequencing</u>

The 10X Genomics protocol ("Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide" available from https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf) was followed to create gel in emulsion beads (GEMs) containing single cells, hydrogel beads and reagents for reverse transcription, perform barcoded cDNA synthesis, and produce sequencing libraries from pooled cDNAs. The concentration of single cell suspensions

was approximately 119 cells/µl (sample 1) and 122 cells/µl (sample 2), as estimated by flow cytometry-based counting (Cytoflex). We concentrated both samples to 732 cells/µl. We loaded two aliquots of sample 1 and two of sample 2 into four 10X reactions according to the 10X protocol (Chromium Single Cell 3' Reagent Kits v3.1), intending to capture approximately 10,000 cells per reaction.

Library construction (following GEM breakage) was using 10X reagents following the "Single Cell 3'Reagent Kits v3.1 User Guide". The four libraries were sequenced on one lane of Illumina Novaseq XP (28 bp read 1, 91 bp read 2, 8 bp index 1, 0 bp index 2). All raw sequence data will be deposited in the ENA.

Mapping and quantification of single-cell RNA-seq

We mapped the single-cell RNA-seq data to version 10 of the *S. mansoni* reference genome (Buddenborg *et al*., (2021) and WormBase ParaSite ((Howe et al., 2017)) using Cell Ranger (version 6.0.1). Prior to mapping, exon 8 was removed from Smp_323380 and the 3' end of was trimmed from Smp_337410 to reduce erroneous mapping to the W chromosome from Z gametologues. The Cell Ranger implementation of EmptyDrops (Lun et al. 2019) was used to detect cells and empty droplets. Across the four samples on average 63.88% of the reads mapped confidently to the transcriptome (63.4% for sorted, 64.35% for unsorted), and an average 83.65% of reads were associated with a cell. There were an average of 118,395 reads per cell. Across the four samples, we captured an estimated 33,319 cells: 8,815 and 24,504 cells from sorted and unsorted samples, respectively. The median gene count was 1,660, and the median UMI count was 5,343. Following QC, 20,478 cells remained for onward analysis.

Quality control (QC) and clustering using Seurat

The four samples were imported into Seurat (version 4.3.0) using R (version 4.1.3) in RStudio (version 1.4.1106) for quality control, clustering, and marker gene identification. The four samples were initially processed separately. Cells with less than 200 genes expressed, and genes in less than three cells were removed. QC metrics including distribution of gene and UMI counts per cell were inspected visually and used to select QC cut-offs. Following a preliminary analysis without further cell filtering, cells with fewer than 500 UMIs were removed and a maximum mitochondrial threshold of 5% was applied. This was based on data distribution, best practice from the authors of the software, and other similar

studies; here we were aiming to remove low-quality cells while being fairly permissive as there may be cell types in these data outside of standard filtering parameters (e.g. expecting anucleate cell type(s)). The data were normalised, scaled (regressing for mitochondrial gene percentage), and variable genes were identified for clustering following Seurat guidelines. Throughout the analysis of these data, several tools were used to select the optimal number of principal components (PCs) to use: Seurat's ElbowPlot and JackStraw, and molecular cross-validation (https://github.com/constantAmateur/MCVR/blob/master/code.R), and Seurat's clustree tool was used to assess the stability of clusters with increasing resolution. We employed DoubletFinder (McGinnis et al., 2019) to identify multiplets, as these can lead to spurious signals. The estimated doublet rate was calculated based on 10X loading guidelines, for our data we estimated multiplet rates at 8% for sorted samples and 52.8% for unsorted samples. We used paramSweep, as recommended, to identify the optimum parameters for each sample in our data, and removed cells classified as multiplets. DoubletFinder removed 299 and 303 cells from the two sorted samples, and 6,069 and 5,496 cells from the two unsorted samples.

We then combined the two sorted samples and the two unsorted samples, and used SCTransform to normalise, and scale these two objects (with regression for mitochondrial percentage and sample ID). Seurat's integrate data function was then used to integrate the sorted and unsorted samples (with normalization.method = "SCT"). This resulted in a total of 20,478 cells passing all the QC filters and going onto the main analysis, 3,849 and 3,962 from the sorted samples and 6,702 and 5,965 from unsorted.

Following integration, principal component analysis (PCA) was run for up to 100 PCs. The data were explored using a range of PCs and resolutions, and using the tools listed above, and 55 PC and a resolution of 1 were selected for the analysis. Top markers were identified using Seurat's FindAllMarkers (min.pct = 0.0, logfc.threshold = 0.0, test.use = "roc") as recommended by Seurat best practices (https://satijalab.org/seurat/) to capture also capture subtle differences in gene expression, and were sorted by AUC. We used the identity and expression of top marker genes (as identified by AUC), along with previously published cell and tissue type marker gene expression to identify the identity of the cell clusters. To attach gene description and chromosome locations to gene IDs, the GFF for the latest publicly available version of the genome was downloaded from WBPS (WBPS17 release, Howe et al., 2017)). The metadata for the genes was isolated and attached to lists of marker genes.

Sex-effect analysis

In order to investigate sex effects on the transcriptome, the cells were categorised as W+ or W- (an approximation of female and male cells respectively, as males lack WSR genes). As ground-truth data was absent, the % of reads mapping to ZSR and WSR was calculated using Seurat's PercentageFeatureSet. Cells with 0.2% or more reads were classified as W+, and the remaining cells were classified as W-. This produced 8,702 cells labelled as W+, and 11,776 cells labelled as W-.

Self-Assembling Manifold (SAM) algorithm

The count data from the QC-filtered Seurat analysis was exported for SAM analysis. For each SAM analysis, the raw count data was imported as a SAM object, standard preprocessing done, and the SAM algorithm run. For further analysis the object was converted into a Scanpy (Wolf et al. (2018)) object, and metadata attached from the Seurat analysis.

EdU-staining experiment

Miracidia were collected as described above for 2 hours. As miracidia continually hatch over this time period, they ranged from 0-2 hours post-hatching. The miracidia were collected and split into two samples; one for EdU labelling and the second for a negative control (miracidia in water without EdU). Sample one was pulsed with 500 µM EdU in water for 4 hours. Sample two was kept in water for 4 hours. Both samples were fixed in 4% PFA for 30 minutes at room temperature. EdU incorporation was detected by click reaction with 25 µM Alexa Fluor488 Azide conjugates (Invitrogen) for 30 minutes. As a positive control we pulsed freshly-transformed mother sporocysts in EdU for 3 days (Wang et al 2013). Three biological replicates were carried, with 20 miracidia and 20 mother sporocysts / treatment examined for each replicate.

In situ hybridisation

Miracidia collected as described above within 4 hours of hatching were fixed in 4% PFA in 2 µm filtered PBSTw (1x PBS + 0.1% Tween) at room temperature for 30 minutes on a rocker. The larvae were transferred to incubation baskets (Intavis, 35 µm mesh) in a 24-well plate and rinsed 3 x 5 minutes in PBSTw, and incubated in 5 µg/ml Proteinase K (Invitrogen) in 1x

PBSTw for 5 minutes at room temperature. They were post-fixed in 4% Formaldehyde in PBSTx for 10 minutes at room temperature then rinsed in PBSTw for 10 minutes. From this point on the *in situ* hybridisation experiments followed the protocol described by Choi *et al*., (2016; 2018) and developed for wholemount nematode larvae. We carried out fluorescent *in situ* hybridisation on 27 cell type marker genes. Probes, buffers, and hairpins for third generation *in situ* hybridization chain reaction (HCR) experiments were purchased from Molecular Instruments (Los Angeles, California, USA). Molecular Instruments designed probes against the sequences on WormBase ParaSite (WBPS version 14, WS271, Howe et al., 2017)(Supplementary table 9). We subsequently labelled some larvae with phalloidin by adding 1 µl phalloidin to 1 ml 5X SSCT for an hour at room temperature in the dark, then rinsing 6 x 30 minutes in SSCT, incubating overnight at 4°C in DAPI fluoromount-G (Southern Biotech) and then mounting and imaging on a confocal laser microscope (Leica Sp8).

Molecular interaction analysis

Molecular interactions were predicted using the online search tool STRING (www.string-db.org; V 11) (Szklarczyk et al 2019) for the following well-defined cell clusters; Neuron 2-6, Protonephridia, Parenchyma 1 & 2, Tegument and the two main stem cell populations - Delta/Phi and Kappa. Protein sequences for the top 100 marker genes for each of these cell clusters were collected from WormBase ParaSite. The protein sequences were entered as a multiple protein search. Default settings were used to predict interactions with either a minimum interaction (confidence) score of 0.4 or 0.7, corresponding to a medium or high level of confidence.

Gene Ontology enrichment analysis

The list of marker genes for each cluster was filtered in order to focus on specific cluster markers. Genes were retained for analysis if they were detected in a minimum 70% of cells of the cluster, less than 30% of cells outside the cluster, and had a minimum AUC score of 0.7. To compare the functional enrichment in Stem 1 v Stem 2, Stem A was combined with Stem B, and Stem C was combined with Stem D, and the DE markers were identified using Seurat's FindMarkers. These markers were filtered (with adjusted p-values <0.001) and used at the basis for GO enrichment analysis in the delta/phi and kappa stem clusters. TopGO (version 2.46.0) was used in R with the weight01 method to identify enriched GO terms.

Node size was set to a minimum of 5 terms, and analysis was run for each cluster on the BP and MF categories. Fisher's exact test was used to assess statistical significance of overrepresented terms, with a FDR threshold set < 0.05.

Funding

The project was funded by the Wellcome Trust through their core support of the Wellcome Trust Sanger Institute (grant 098051 and 206194).

Author Contribution

KR, TA, KA, GR and MB designed the research

CD developed and optimised dissociation protocol

JG carried out cytometry

GS optimised EdU protocol

TA analyzed data

KR performed ISH experiments

SD annotated marker genes in v10 of the genome

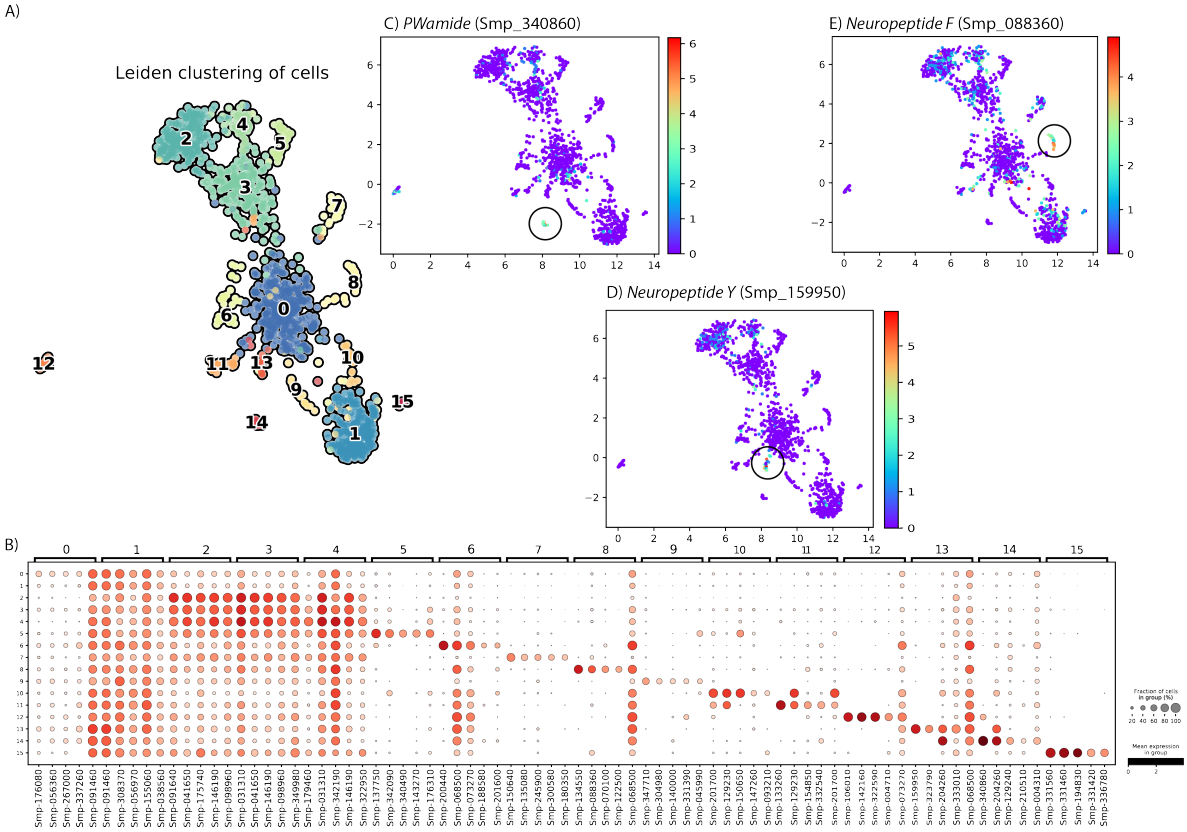
TA and KR drafted the complete manuscript, with sections of text contributed by GR and MB.

All authors read and approved the final manuscript.

Acknowledgements

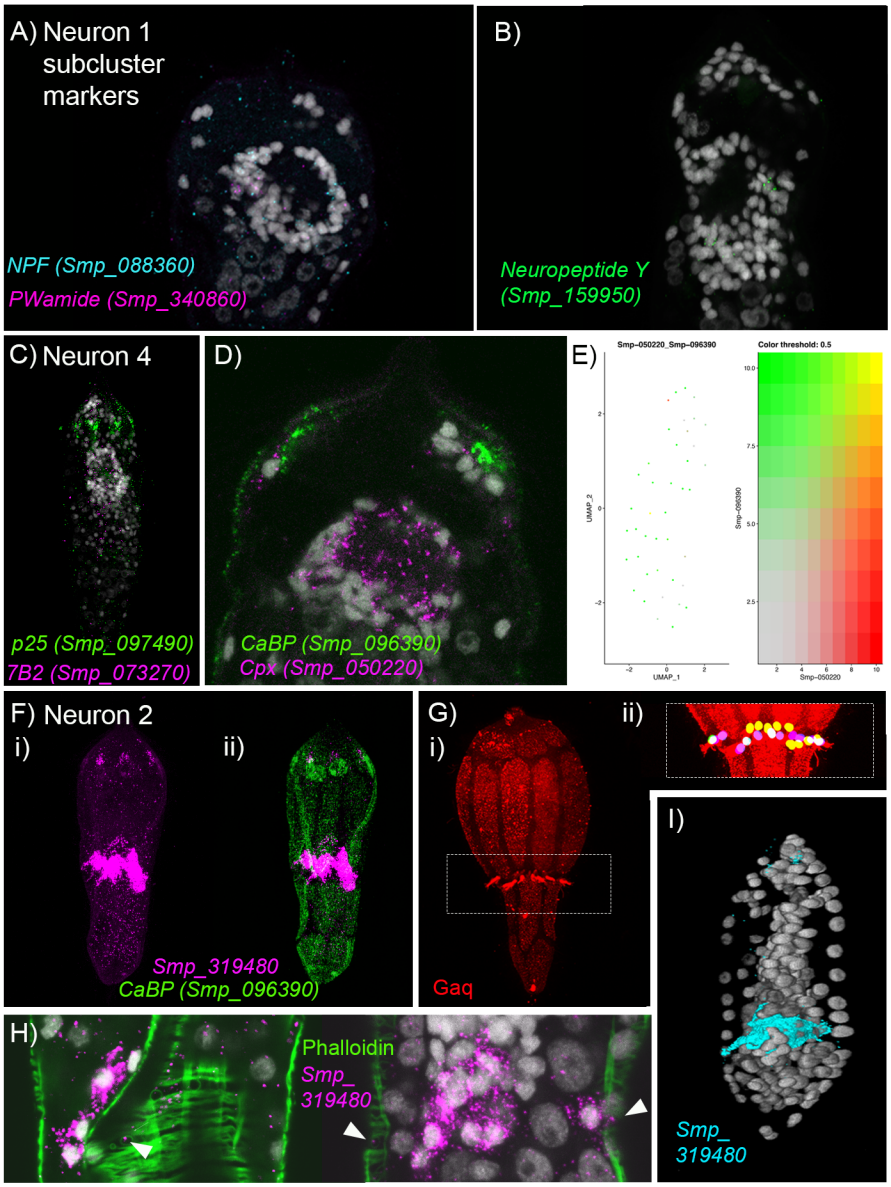
We thank Simon Clare and his team for technical assistance with the life cycle maintenance and collection of parasite material; also Karl Hoffmann, Josephine Fforde-Thomas and Benjamin Hulme for on-going parasite material. We thank Sarah Buddenborg for work on the annotation of the *Schistosoma mansoni* genome; David Goulding for assistance with Microscopy; teams within DNA Pipelines for data production; and core Informatics teams and Pathogen Informatics for the infrastructure used for data analysis. We thank members of the Parasite Genomics team at the Wellcome Sanger Institute for their comments and input on this study.

Supplementary figure legends



Supplementary figure 1. Subclustering the Neuron 1 cluster using SAM revealed multiple distinct subpopulations.

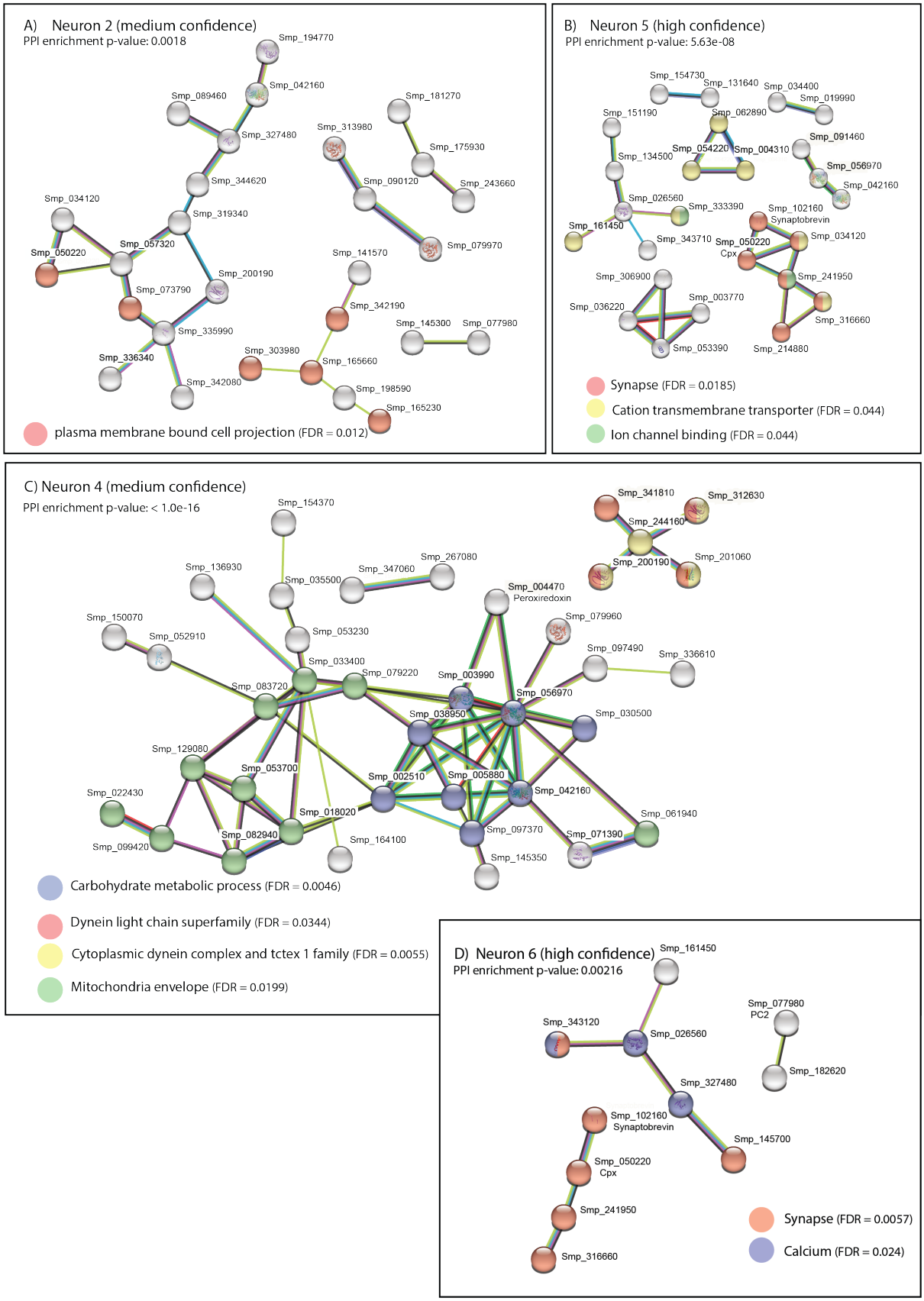
A) The Neuron 1 cluster was analysed using Self-Assembling Manifold (SAM) algorithm, and Leiden clustering indicated 16 subpopulations. **B)** The top 5 marker genes for each of the subpopulations. **C)** Expression of NPP PWamide, the top marker for subcluster 14. **D)** Expression of neuropeptide Y, the top marker for subcluster 13. **E)** Expression of neuropeptide F, the second top marker for subcluster 8. Although this cluster was difficult to define, this analysis suggests that there may be several small neural subpopulations within this cluster.



Supplementary figure 2. Neurons

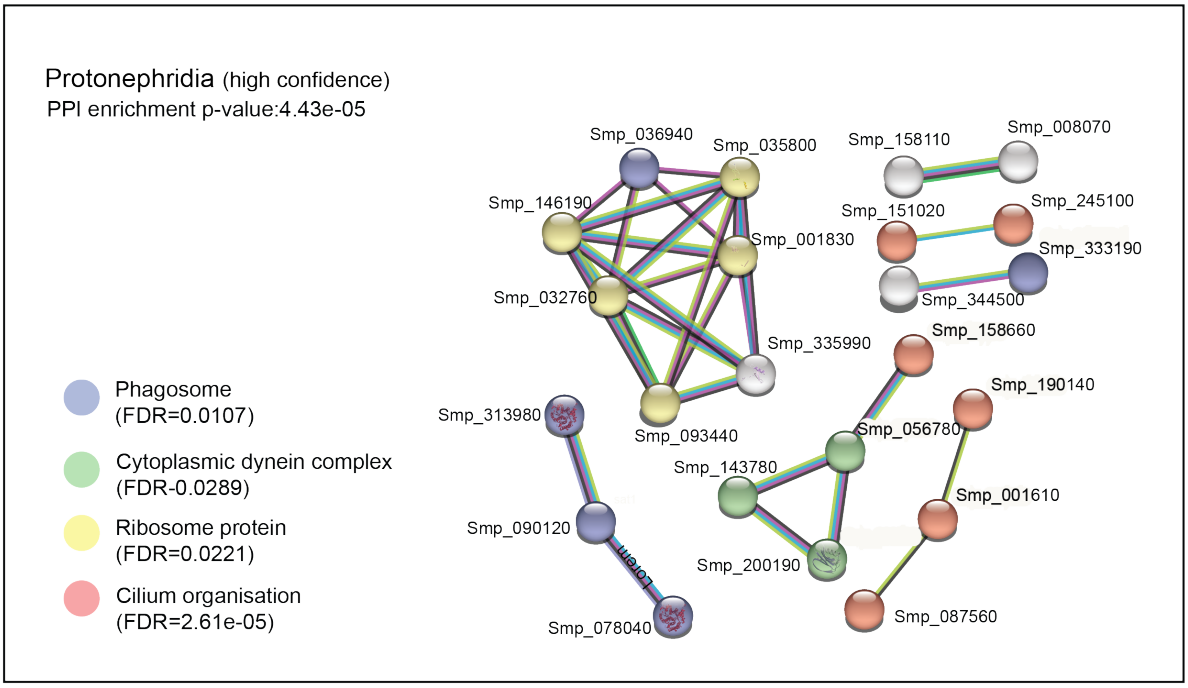
A & B) Top markers for Neuron 1 include the neuropeptides precursor genes for Neuropeptide F, PWamide, and Neuropeptide Y. ISH of these genes shows expression in cells that make up the brain. **C)** The expression of neuron 4 marker *p25 alpha* (which is a marker for ciliated cells in adult worms) suggests that the submuscular cells maybe ciliated. **D & E)** To investigate if neuron 4 cells are ciliated plates or the submuscular cells we looked at co-expression of *cpx* and *CaBP in situ* **(D)** and *in silico* **(E).** *In situ* expression shows co-expression of *cpx* and *CaBP* in the submuscular cells but not the ciliary plates. *In silico* expression shows only 1 cell with strong co-expression, this may be due to gene drop out of *cpx* or it could suggest that most neuron 4 cells are ciliary plates. **F)** The top marker for neuron 2 *Smp_319480* (putative mTORC1) is expressed in **i)** 4 cells anteriorly and ~22 that sit lateral and just posterior to the brain, sending projections into the brain and out to the body wall at the latitude between the 2nd and 3rd tier of ciliary plates **(ii)** seen here expressing the neuron 4 marker gene *Smp_096390* (arrowheads). **G) i & ii)** G alpha-q antibody staining reveals a girdle of 23 multi-ciliated nerve endings that sit between the 2nd and 3rd tier of

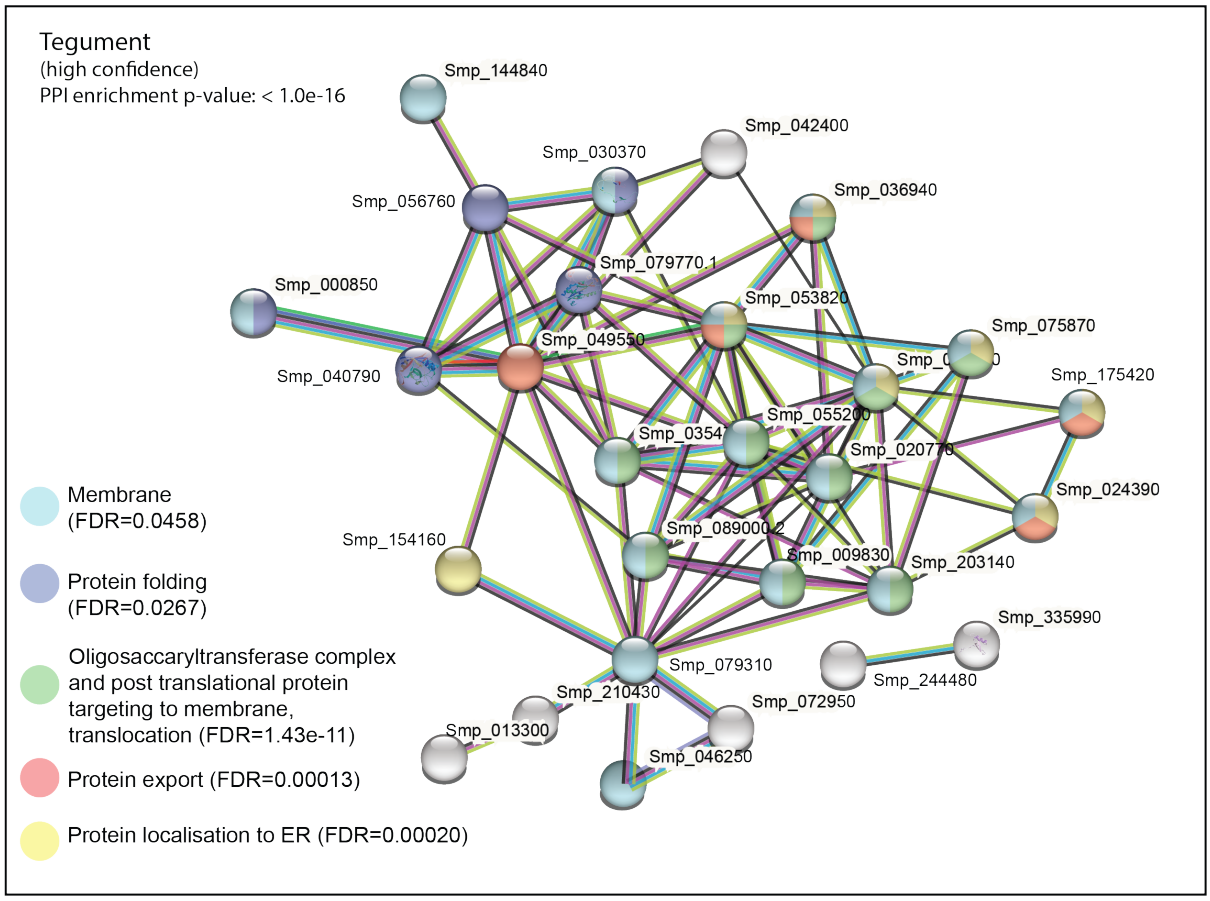
ciliary plates. **H)** Neuron 2 marker Smp_319480 transcripts are expressed at the base of these cilia; **I)** and some transcripts also reach the nuclei of circular body wall muscles.



Supplementary figure 3. STRING analysis of neural clusters.
STRING analysis of the top 100 marker genes for each of the neural populations whose predicted networks have a functional enrichment. **A)** Neuron 2, **B)** Neuron 5, **C)** Neuron 4,

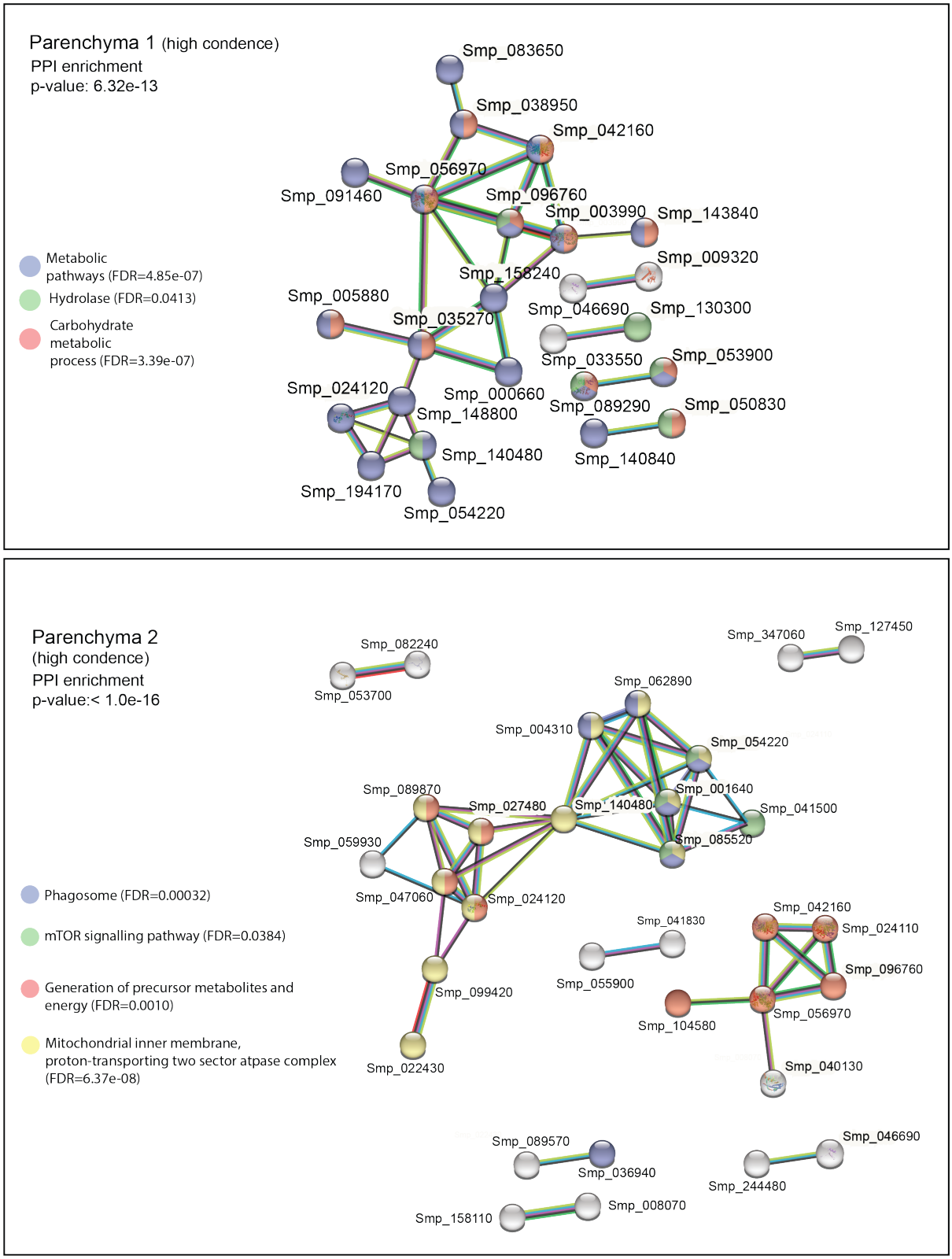
D) Neuron 6. Neuron 1 is not a defined cluster and Neuron 3 has no functional enrichment. Lines (edges) connecting nodes are based on evidence of the function of homologues. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum interaction (confidence) score of 0.7, corresponding to a high level of confidence, 0.4 medium level of confidence.





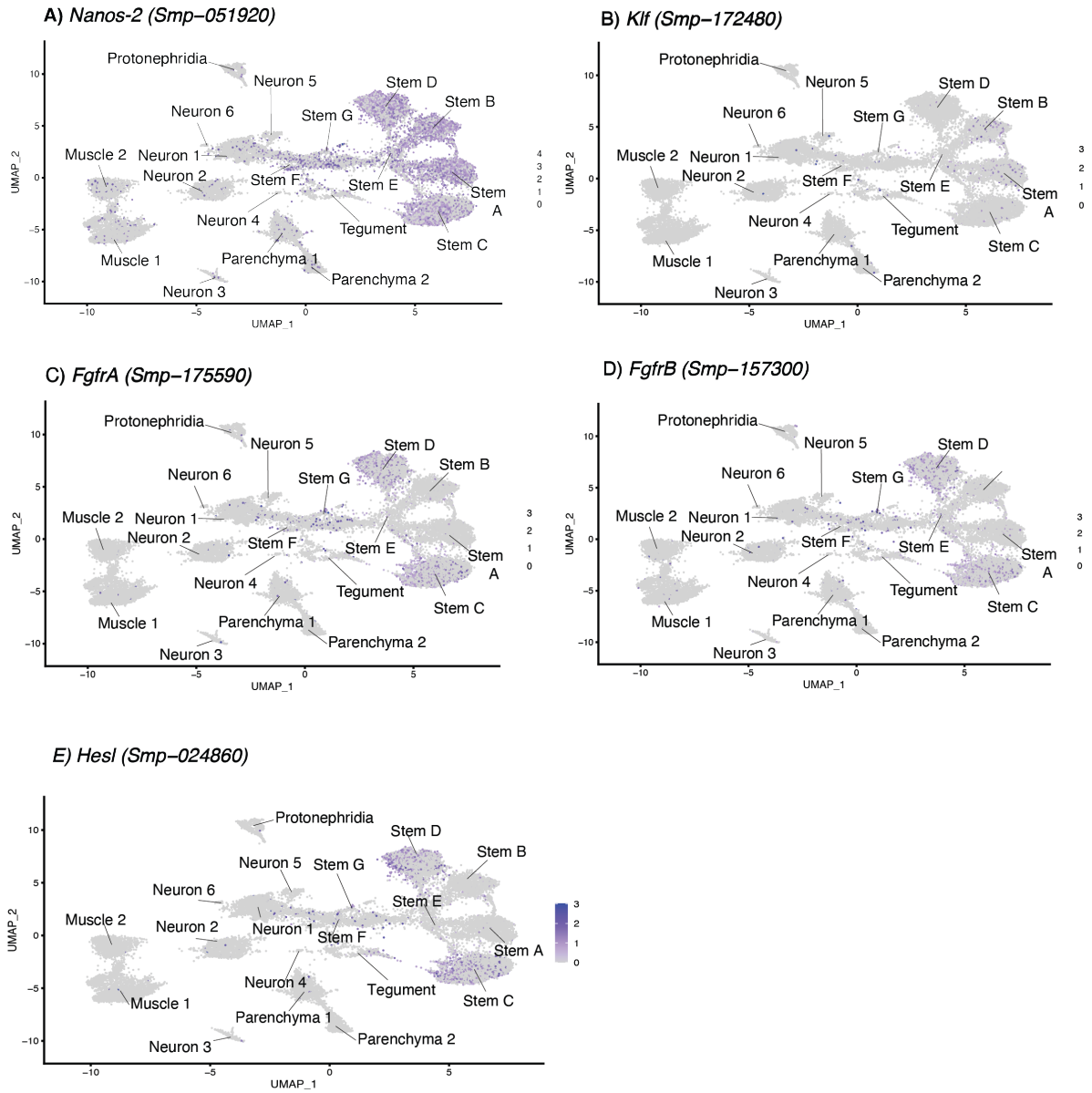
Supplementary figure 4. STRING analysis of protonephridia and tegument clusters
STRING analysis of the top 100 marker genes for **A)** protonephridia, **B)** tegument. Lines (edges) connecting nodes are based on evidence of the function of homologues. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum

interaction (confidence) score of 0.7, corresponding to a high level of confidence, 0.4 medium level of confidence.

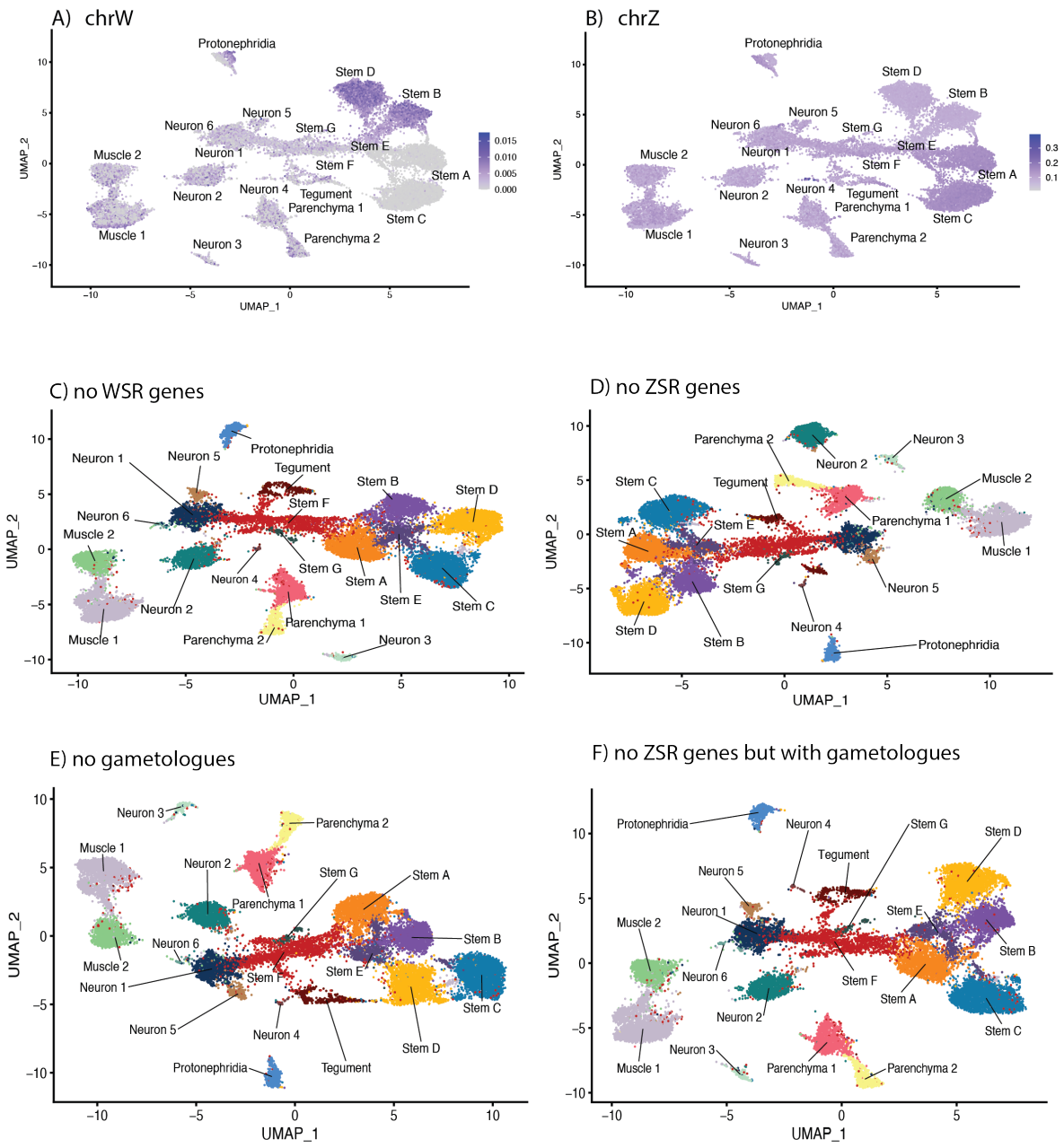


Supplementary figure 5. STRING analysis of parenchyma clusters

STRING analysis of the top 100 marker genes for **A)** parenchyma 1, **B)** parenchyma 2. Lines (edges) connecting nodes are based on evidence of the function of homologues. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum interaction (confidence) score of 0.7, corresponding to a high level of confidence, 0.4 medium level of confidence.

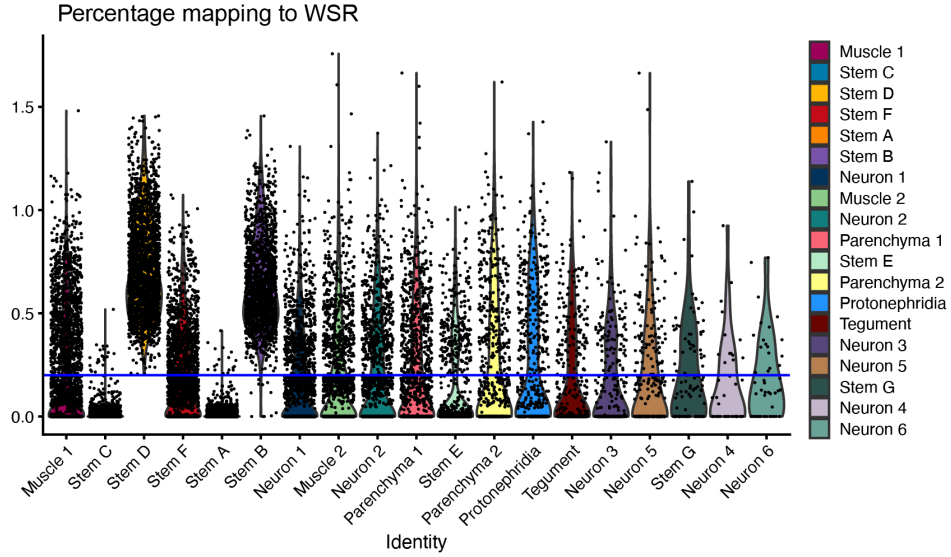


Supplementary figure 6. Expression of known stem cell markers in miracidia cells
A-E) Three stem cell classes were identified in the mother sporocyst stage and were named based upon their respective markers: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*nanos-2*$^+$ and *fgfrA*$^+$, *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$and, *fgfrB*$^+$ and *hesl*$^+$)(Wang *et al.*, 2018; Safarti *et al.*, 2021).
In the miracidium, we identified seven stem cell clusters by expression of *ago2-1*: stem A and B are *klf*$^+$ and *nanos-2*$^+$ (i.e. Kappa-like), stem C and D are *nanos-2*$^+$, *fgfrA*$^+$, *fgfrB*$^+$,and *hesl*$^+$ (i.e. they resembled both Delta and Phi).

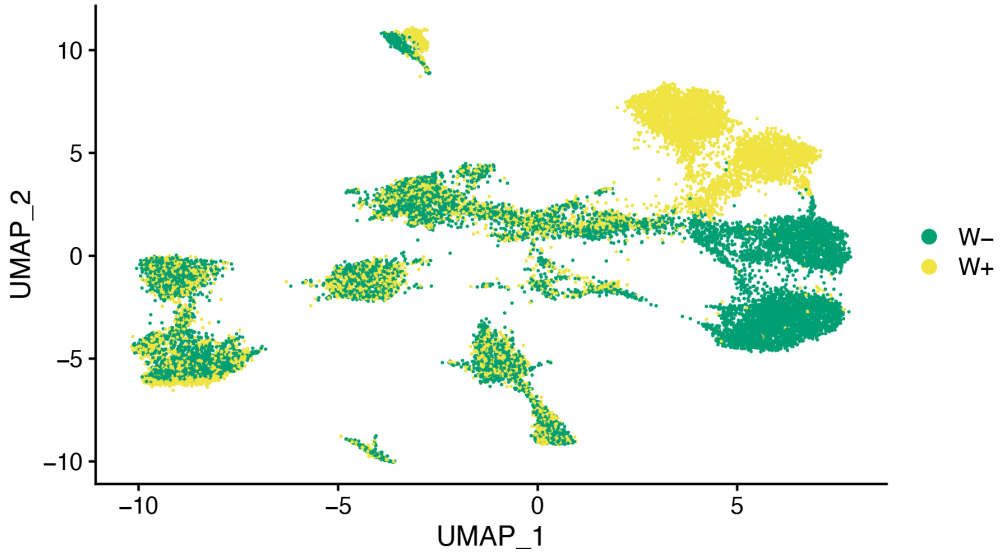


Supplementary figure 7. The effect of sex-linked genes on cell clustering.
A & B) The proportion of reads mapping to the WSR and ZSR across all cell type clusters shows a) putative male cells (i.e. those that lack WSR gene expression) and b) evidence of incomplete ZSR gene dosage compensation in the male stem cells. **C)** To determine whether the 29 WSR genes were responsible for the clustering pattern, we excluded them from the mapping reference. The impact on the overall UMAP clustering was minimal. We repeated this excluding the **D)** ZSR genes, **E)** gametologues and **F)** ZSR genes except the gametologues in turn.

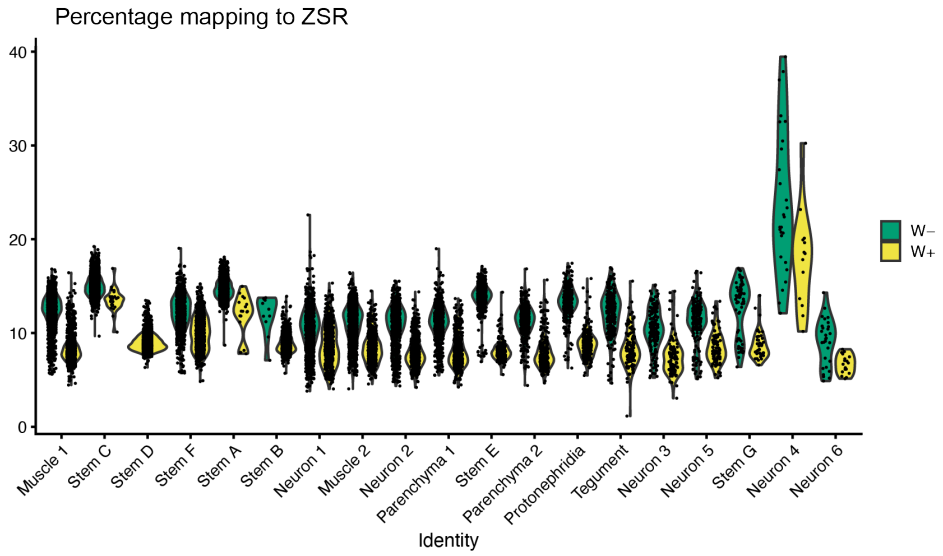
A)


Percentage mapping to WSR
1.5
1.0
0.5
0.0
Muscle 1
Stem C
Stem D
Stem F
Stem A
Stem B
Neuron 1
Muscle 2
Neuron 2
Parenchyma 1
Stem E
Parenchyma 2
Protonephridia
Tegument
Neuron 3
Neuron 5
Stem G
Neuron 4
Neuron 6
Identity

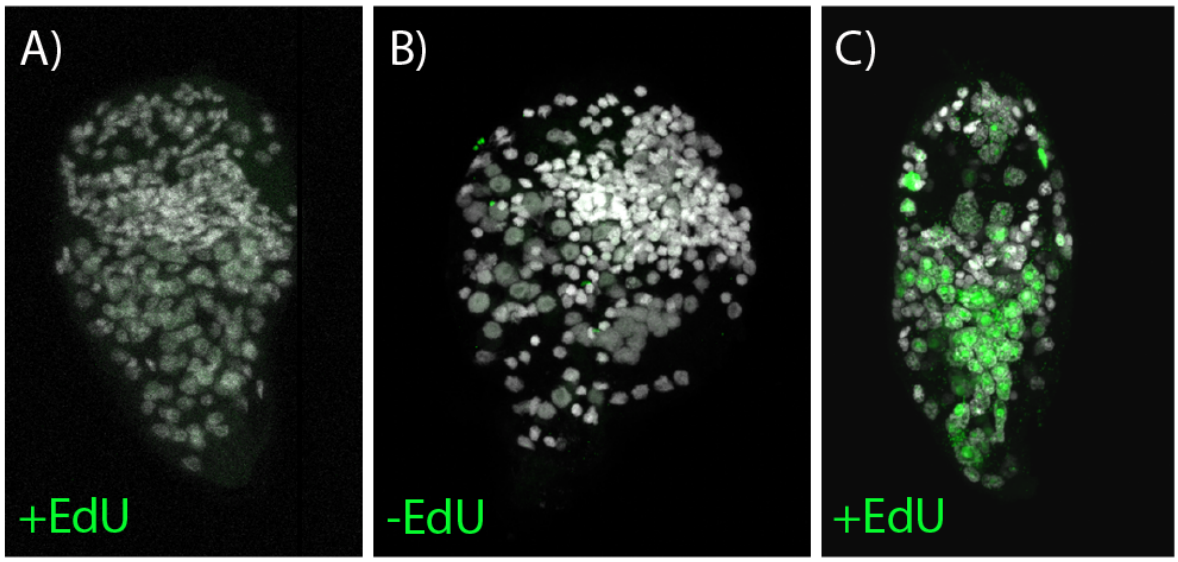
B)


10
5
0
−5
−10
UMAP_2
−10
−5
0
5
UMAP_1
W−
W+

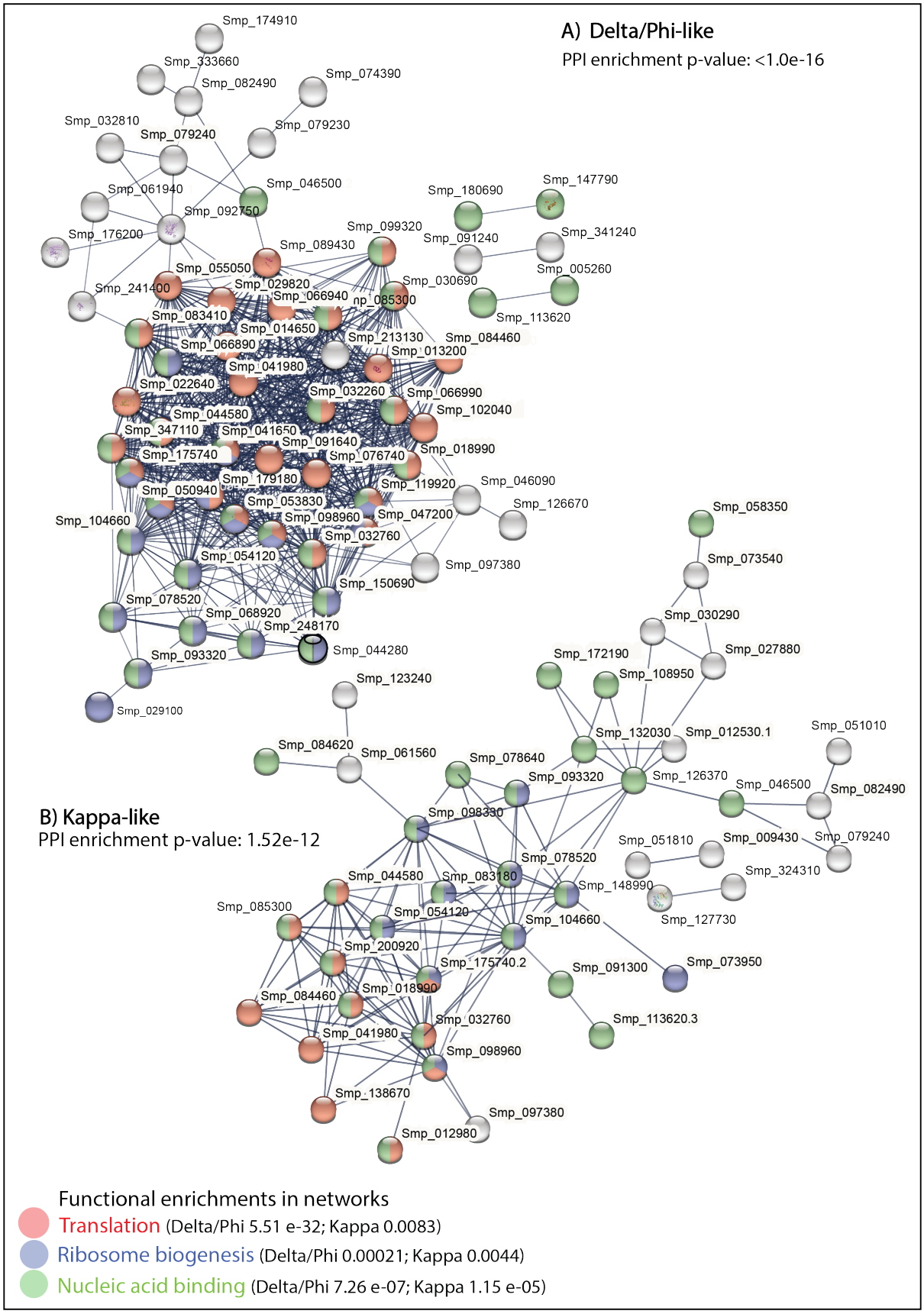
C)


Percentage mapping to ZSR
40
30
20
10
0
Muscle 1
Stem C
Stem D
Stem F
Stem A
Stem B
Neuron 1
Muscle 2
Neuron 2
Parenchyma 1
Stem E
Parenchyma 2
Protonephridia
Tegument
Neuron 3
Neuron 5
Stem G
Neuron 4
Neuron 6
Identity
W−
W+

Supplementary figure 8. Pseudotyping of cells based on percentage of reads mapping to WSR indicates incomplete dosage compensation of ZSR genes.
A) The percentage of reads mapping to W specific genes by cluster. The blue line at 0.2% indicates where the data were split into W+ and W- categories, i.e. stem B and D contain cells with reads mapping to W specific genes (and are likely cells from female miracidia), whereas most cells in stem A and C do not map to W specific genes (cells from male miracidia). **B)** UMAP showing the clusters by W+ and W- categories. Almost complete separation can be seen in the main stem cell clusters, and the categories are mixed in most other clusters. **C)** The percentage of reads mapping to Z specific genes by cluster, separated by W+/W-. Differences in ZSR mapping percentage between the categories are pronounced in Stem A to D but can also be observed across the clusters. These differences between W+ and W- cells within each cluster suggest that there may be incomplete dosage compensation across most clusters.

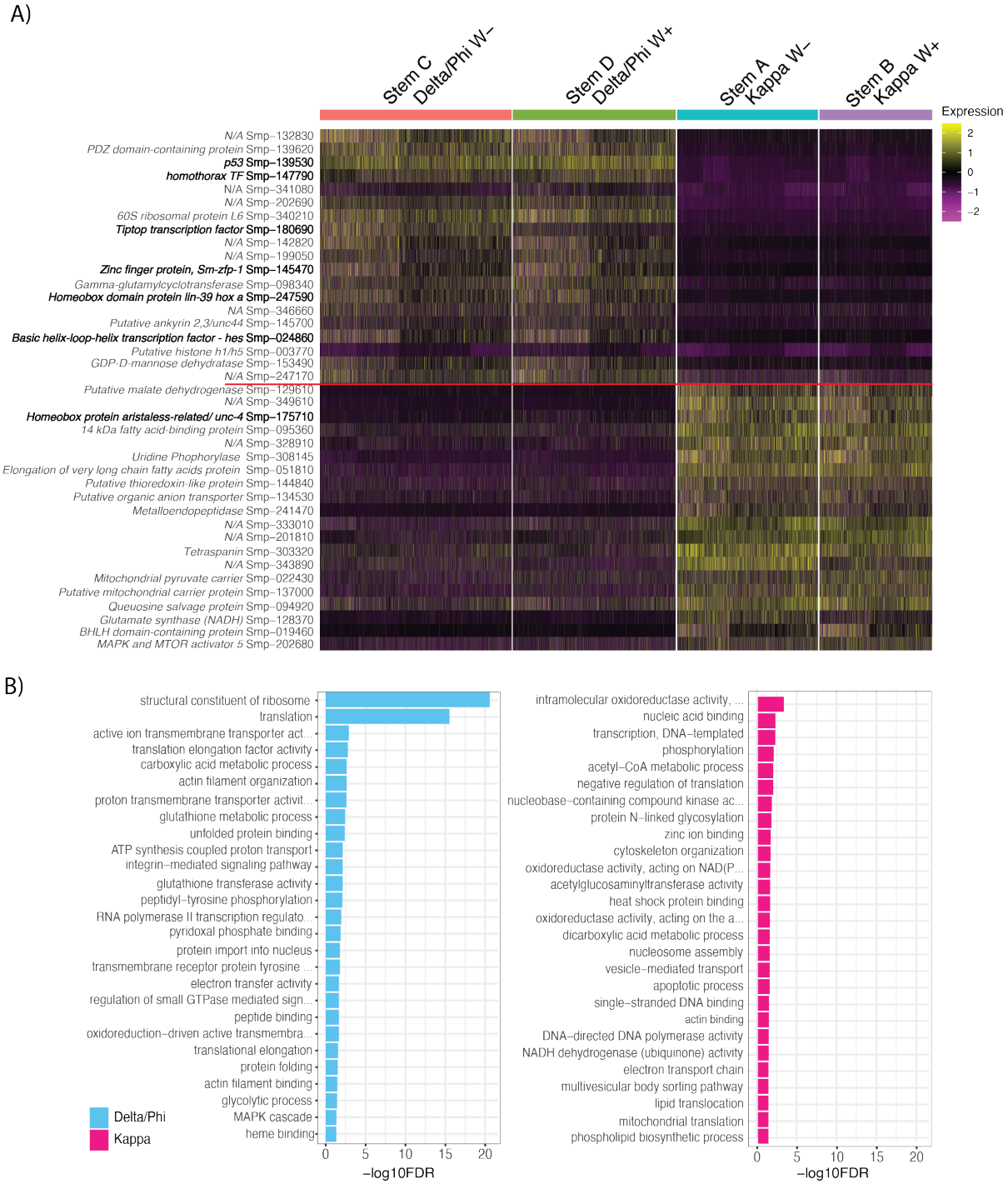


Supplementary figure 9. EdU labelling shows no proliferating cells in miracidia.
A) 4 hour EdU pulse on miracidia 0-6 hours post-hatching shows no EdU incorporation in any cells. **B)** negative control. **C)** Positive control, 3 day EdU pulse on freshly-transformed mother sporocysts shows many dividing cells incorporating EdU, all treatments n = 30.



Supplementary figure 10. STRING analysis of the top 100 marker genes for the two stem cell populations in the miracidium (combining the female and male cells).

A) 61 of the top markers of Stem C and D combined (i.e. Delta/Phi-like) and **B)** 37 for Stem A and B combined (i.e. Kappa-like) form large predicted interactions enriched for translation, ribosome biogenesis and nucleic acid binding. Lines (edges) connecting nodes are based on evidence of the function of homologues. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum interaction (confidence) score of 0.7, corresponding to a high level of confidence.



Supplementary figure 11. Differential gene expression between the two stem cell populations; Delta-phi and Kappa.
A) Heatmap of top 20 differentially expressed genes show high number of transcription factors (bold) in Delta/Phi and enrichment of lipid and glycolytic metabolism genes in Kappa. **B)** Gene ontology analysis of DEG (adj $p<0.001$) revealed that upregulated genes in Delta/Phi were related to structural constituent of the ribosome and translation, while up-regulated genes in Kappa were related to transcription and metabolism.

References

Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Development. 2017 Oct 1;144(19):3625-3632. doi: 10.1242/dev.151142. Epub 2017 Aug 29. PMID: 28851704; PMCID: PMC5665481.

Ashton PD, Harrop R, Shah B, Wilson RA. The schistosome egg: development and secretions. Parasitology. 2001 Mar;122(3):329-38.

Bezares-Calderón LA, Berger J, Jasek S, Verasztó C, Mendes S, Gühmann M, Almeda R, Shahidi R, Jékely G. Neural circuitry of a polycystin-mediated hydrodynamic startle response for predator avoidance. Elife. 2018 Dec 14;7:e36262. doi: 10.7554/eLife.36262. PMID: 30547885; PMCID: PMC6294549.

Buddenborg SK, Tracey A, Berger DJ, Lu Z, Doyle SR, Fu B, Yang F, Reid AJ, Rodgers FH, Rinaldi G, Sankaranarayanan G. Assembled chromosomes of the blood fluke Schistosoma mansoni provide insight into the evolution of its ZW sex-determination system. bioRxiv. 2021 Aug 13:2021-08.

Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, et al. (2012) TrakEM2 Software for Neural Circuit Reconstruction. PLoS ONE 7(6): e38011. https://doi.org/10.1371/journal.pone.0038011

Choi HM, Calvert CR, Husain N, Huss D, Barsi JC, Deverman BE, Hunter RC, Kato M, Lee SM, Abelin AC, Rosenthal AZ. Mapping a multiplexed zoo of mRNA expression. Development. 2016 Oct 1;143(19):3632-7.

Choi HM, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, Pierce NA. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development. 2018 Jun 15;145(12):dev165753.

Chernin E. Interference with the capacity of Schistosoma mansoni miracidia to infect the molluscan host. The Journal of Parasitology. 1968 Jun 1:509-16.

Collins III JJ, King RS, Cogswell A, Williams DL, Newmark PA. An atlas for Schistosoma mansoni organs and life-cycle stages using cell type-specific markers and confocal microscopy. PLoS neglected tropical diseases. 2011 Mar 8;5(3):e1009.

Costain AH, MacDonald AS, Smits HH. (2018) Schistosome Egg Migration: Mechanisms, Pathogenesis and Host Immune Responses. Front Immunol. 9:3042. doi: 10.3389/fimmu.2018.03042. Erratum in: Front Immunol. 2019 Apr 11;10:749.

Diaz Soria CL, Lee J, Chong T, Coghlan A, Tracey A, Young MD, Andrews T, Hall C, Ng BL, Rawlinson K, Doyle SR et al. 2020. Single-cell atlas of the first intra-mammalian developmental stage of the human parasite *Schistosoma mansoni*. Nature Communications;11(1):6411.

Gąsiorowski L, Andrikou C, Janssen R, Bump P, Budd GE, Lowe CJ, Hejnol A, Molecular evidence for a single origin of ultrafiltration-based excretory organs, Current Biology, 2021, 31,16;3629-3638.https://doi.org/10.1016/j.cub.2021.05.057.

Kevin L. Howe, Bruce J. Bolt, Myriam Shafie, Paul Kersey, and Matthew Berriman. WormBase ParaSite − a comprehensive resource for helminth genomics *Molecular and Biochemical Parasitology* 2017 215 2-10 10.1016/j.molbiopara.2016.11.005

Kevin L. Howe, Bruce J. Bolt, Scott Cain, Juancarlos Chan, Wen J. Chen, Paul Davis, James Done, Thomas Down, Sibyl Gao, Christian Grove, Todd W. Harris, Ranjana Kishore, Raymond Lee, Jane Lomax, Yuling Li, Hans-Michael Muller, Cecilia Nakamura, Paulo Nuin, Michael Paulini, Daniela Raciti, Gary Schindelman, Eleanor Stanley, Mary Ann Tuli, Kimberly Van Auken, Daniel Wang, Xiaodong Wang, Gary Williams, Adam Wright, Karen Yook, Matthew Berriman, Paul Kersey, Tim Schedl, Lincoln Stein, and Paul W. Sternberg. WormBase 2016: expanding to enable helminth genomic research
Nucleic Acids Research 2016 44(D1) D774-D780
10.1093/nar/gkv1217

Jurberg, A.D., Gonçalves, T., Costa, T.A. et al. The embryonic development of Schistosoma mansoni eggs: proposal for a new staging system. Dev Genes Evol 219, 219–234 (2009). https://doi.org/10.1007/s00427-009-0285-9

Kakugawa, S., Langton, P., Zebisch, M. et al. Notum deacylates Wnt proteins to suppress signalling activity. Nature 519, 187–192 (2015). https://doi.org/10.1038/nature14259

Kharchenko, P., Silberstein, L. & Scadden, D. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11, 740–742 (2014). https://doi.org/10.1038/nmeth.2967

Kumar P, Tan Y, Cahan P. Understanding development and stem cells using single cell-based analyses of gene expression. Development. 2017 Jan 1;144(1):17-32. doi: 10.1242/dev.133058.

Li P, Nanes Sarfati D, Xue Y, Yu X, Tarashansky AJ, Quake SR, Wang B. Single-cell analysis of Schistosoma mansoni identifies a conserved genetic program controlling germline stem cell fate. Nature Communications. 2020 Jan 20;12(1):485.

Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba A. Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. Cell Stem Cell. 2015 Sep 3;17(3):329-40. doi: 10.1016/j.stem.2015.07.002. Epub 2015 Jul 30. PMID: 26235341.

Mair, G., Maule, A., Day, T., Halton, D. (2000). A confocal microscopical study of the musculature of adult Schistosoma mansoni. Parasitology, 121(2), 163-170. doi:10.1017/S0031182099006174

Maldonado, José F., and Josefina Acosta-Matienzo. "Biological studies on the miracidium of Schistosoma mansoni." *The American journal of tropical medicine and hygiene* 1.5 (1948): 645-657.

Mann VH Morales ME Rinaldi G Brindley PJ (2010) Culture for genetic manipulation of developmental stages of Schistosoma mansoni Parasitology 137:451–462.
McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou XN. Schistosomiasis. Nat Rev Dis Primers. 2018 Aug 9;4(1):13. doi: 10.1038/s41572-018-0013-8. PMID: 30093684.

McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019 Apr 24;8(4):329-337.e4. doi: 10.1016/j.cels.2019.03.003. Epub 2019 Apr 3. PMID: 30954475; PMCID: PMC6853612.

McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou XN. Schistosomiasis. Nat Rev Dis Primers. 2018;4(1):13. doi: 10.1038/s41572-018-0013-8. PubMed PMID: 30093684.

Nanes Sarfati D, Li P, Tarashansky AJ, Wang B. Single-cell deconstruction of stem-cell-driven schistosome development. Trends in Parasitology. 2021 Sep;37(9):790-802. DOI: 10.1016/j.pt.2021.03.005. PMID: 33893056; PMCID: PMC8364473.

Orosz F, Ovádi J. TPPP orthologs are ciliary proteins. FEBS Lett. 2008 Nov 12;582(27):3757-64. doi: 10.1016/j.febslet.2008.10.011. Epub 2008 Oct 16. PMID: 18930059.

Pan SC (1980) The fine structure of the miracidium of Schistosoma mansoni Journal of Invertebrate Pathology 36:307–372.

Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body axis. Cell. 2009 Dec 11;139(6):1056-68. doi: 10.1016/j.cell.2009.11.035. PMID: 20005801.

Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, Grunau C, Boissier J, Cosseau C. Dosage Compensation throughout the Schistosoma mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome. Genome Biol Evol. 2019 Jul 1;11(7):1909-1922. doi: 10.1093/gbe/evz133.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019 Jan 8;47(D1):D607-13.

Scimone ML, Cote LE, Reddien PW. Orthogonal muscle fibres have different instructive roles in planarian regeneration. Nature. 2017 Nov 30;551(7682):623-628. doi: 10.1038/nature24660. Epub 2017 Nov 22. PMID: 29168507; PMCID: PMC6263039.

Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. Stem cell differentiation trajectories in *Hydra*resolved at single-cell resolution. Science. 2019 Jul 26;365(6451):eaav9314. doi: 10.1126/science.aav9314.

Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming GL, Song H. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell. 2015 Sep 3;17(3):360-72. doi: 10.1016/j.stem.2015.07.013.

Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Developmental biology. 1977 Mar 1;56(1):110-56.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug 25;126(4):663-76. doi: 10.1016/j.cell.2006.07.024. Epub 2006 Aug 10. PMID: 16904174.

Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. Elife 8, (2019).

Valsecchi CIK, Basilicata MF, Semplicio G, Georgiev P, Gutierrez NM, Akhtar A. Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. Nat Commun. 2018 Sep 7;9(1):3626. doi: 10.1038/s41467-018-05642-2. PMID: 30194291; PMCID: PMC6128902.

Verasztó C, Jasek S, Gühmann M, Shahidi R, Ueda N, Beard JD, Mendes S, Heinz K, Bezares-Calderón LA, Williams E, Jékely G (2020) Whole-animal connectome and cell-type complement of the three-segmented Platynereis dumerilii larva BioRxiv, 2020.08. 21.260984

Vergara HM, Pape C, Meechan KI, Zinchenko V, Genoud C, Wanner AA, Mutemi KN, Titze B, Templin RM, Bertucci PY, Simakov O. Whole-body integration of gene expression and single-cell morphology. Cell. 2021 Sep 2;184(18):4819-37.

Vogg MC, Owlarn S, Rico YA, Xie J, Suzuki Y, Gentile L, Wu W, Bartscherer K. Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. Developmental Biology. 2014 Jun 15;390(2):136-48.

Wang B Collins JJ Newmark PA (2013) Functional genomic characterization of neoblast-like stem cells in larval Schistosoma mansoni eLife 2:e00768.

Wang T, Zhao M, Rotgans BA, Strong A, Liang D, Ni G, Limpanont Y, Ramasoota P, McManus DP, Cummins SF. Proteomic Analysis of the Schistosoma mansoni Miracidium. PLoS One. 2016 Jan 22;11(1):e0147247. doi: 10.1371/journal.pone.0147247.

Wang B, Lee J, Li P, Saberi A, Yang H, Liu C, Zhao M, Newmark PA. Stem cell heterogeneity drives the parasitic life cycle of Schistosoma mansoni. Elife. 2018 Jul 10;7:e35449.

Wendt GR Collins JN Pei J Pearson MS Bennett HM Loukas A Berriman M Grishin NV Collins JJ (2018) Flatworm-specific transcriptional regulators promote the specification of tegumental progenitors in Schistosoma mansoni eLife 7:e33221.

Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, Reese ML, Collins III JJ. A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood feeding. Science. 2020 Sep 25;369(6511):1644-9.

Wendt GR, Shiroor DA, Adler CE, Collins III JJ. Convergent evolution of a genotoxic stress response in a parasite-specific p53 homolog. Proceedings of the National Academy of Sciences. 2022 Sep 13;119(37):e2205201119.

Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW. Muscle cells provide instructions for planarian regeneration. Cell reports. 2013 Aug 29;4(4):633-41.

Supplementary video legends

Supplementary video 1 – Confocal z-stack of wholemount *in situ* hybridisation (WISH) of pan-muscle marker *myosin heavy chain* (*MYH* Smp_085540) (cyan), counterstained with phalloidin (green) and DAPI (white), reveals which *MYH*$^{+}$ nuclei belong to the circular and longitudinal body wall muscles.

Supplementary video 2 – Segmentation of *myosin heavy chain* (Smp_085540) (cyan) positive cells (using the z-stack from supplementary figure 1); the nuclei of the circular body wall muscles are highlighted in orange, the nuclei of the longitudinal body wall muscles are highlighted in yellow.

Supplementary video 3 - Confocal z-stack of WISH of muscle cluster 1 markers *papillin* (Smp_052230) (cyan) and *Wnt-11*-(Smp_156540) (magenta), counterstained with phalloidin (green) and DAPI (white), show expression in circular body wall muscles.

Supplementary video 4 – Maximum intensity projection from confocal z-stack of WISH of muscle cluster 1 markers *notum* (Smp_015820) (cyan) and *Wnt-11*-(Smp_156540) (magenta) counterstained with phalloidin (green) showing expression at opposite poles.

Supplementary video 5 - Confocal z-stack of WISH of muscle cluster 2 markers *calcium activated potassium channel* (Smp_156150) (cyan) and *myoD*

(Smp_167400) (magenta), counterstained with phalloidin (green) and DAPI (white), shows expression in longitudinal body wall muscles.

Supplementary video 6 - Confocal z-stack of WISH of markers for neuron cluster 3 (Smp_201600, magenta) and neuron 6 (Smp_071050, cyan), counterstained with DAPI (white), shows expression of both genes in cells whose nuclei form part of the nuclear rind of the brain.

Supplementary video 7 - Confocal z-stack of WISH of neuron 4 marker *16 kDa calcium binding protein*(Smp_096390), counterstained with phalloidin (green) and DAPI (white), shows expression in the ciliary plates and six previously undescribed cells that we term sub-muscular cells. The nuclei of the sub-muscular cells sit beneath the body wall muscles, each cell has a protrusion (possible cilia) that extends beyond the body wall muscles, between the first and second tiers of ciliary plates, and out to the exterior of the larva.

Supplementary video 8 – Maximum intensity projection of confocal z-stack of WISH of neuron 4 marker *16 kDa calcium binding protein* (Smp_096390), counterstained with phalloidin (green) and DAPI (white), shows expression in ciliary plates and sub-muscular cells.

Supplementary video 9 – Maximum intensity projection from confocal z-stack of WISH of neuron 2 marker Smp_319480 (magenta), counterstained with DAPI (white), showing expression in clusters of cells posterior to the brain that extend into the brain and out to the periphery.

Supplementary video 10 – Maximum intensity projection from confocal z-stack of WISH of neuron 2 marker Smp_319480 (magenta) and neuron 4 marker *16 kDa calcium binding protein* (Smp_096390) (green), counterstained with DAPI (white), showing that the clusters of neuron 2 cells reach the periphery at the latitude between the second and third tiers of ciliary plates.

Supplementary video 11 - Maximum intensity projection from confocal z-stack of WISH of neuron 5 marker Smp_319480 (cyan), counterstained with DAPI (white), showing expression

in cells anterior to the brain, and in and around the paired lateral glands and their secretory ducts.

Supplementary video 12 - Maximum intensity projection from confocal z-stack of WISH of protonephridia marker *sialidase* (Smp_335600) (magenta), counterstained with DAPI (white). This shows the transcripts in a bilaterally symmetrical pattern distributed along the s-shaped path of an excretory tubule that connects the anterior and posterior flame cells.

Supplementary video 13 - Confocal z-stack of WISH of markers for parenchyma 1 and 2 (Smp_318890, green) and stem cells *ago2-1* (Smp_179320, magenta), counterstained with DAPI (white), shows the long cytoplasmic protrusions of the parenchymal cells reaching between the stem and other cells.

Supplementary video 14 - Maximum intensity projection from confocal z-stack of WISH of marker for parenchyma 1 and 2 (Smp_318890, green) counterstained with DAPI (white), shows the distribution of the parenchymal cells and their reach in the intercellular space of the larva.

Supplementary video 15 - Confocal z-stack of WISH of markers for tegument *Meg6* (Smp_163710, magenta) and *dynein light chain* (Smp_200180) (cyan), counterstained with phalloidin (green) and DAPI (white), shows the nuclei sit below the body wall muscle, and cytoplasmic protrusions reach between muscle filaments and form the epidermal ridges between the ciliary plates.

Supplementary video 16 - Confocal z-stack of WISH of markers for stem clusters Delta/Phi *p53* (Smp_139530, cyan), Kappa *UPPA* (Smp_308145, green) and stem E *Phosphomevalonate kinase* (Smp_072460, magenta),counterstained with DAPI (white), shows there are no cells that are *pmk*-positive but negative for *UPPA* and *p53*. Stem E cells cannot be validated *in situ*.

Supplementary video 17 - Confocal z-stack of WISH of markers for stem clusters Delta/Phi *p53* (Smp_139530, cyan), Kappa *UPPA* (Smp_308145, green) and the pan-stem marker *ago2-1* (Smp_179320), counterstained with DAPI (white), shows there are no cells that are *ago2-1* positive but negative for *UPPA* and *p53* that could be Stem F and G cells.