A single-cell atlas of the miracidium larva of the human blood fluke *Schistosoma mansoni*: cell types, developmental pathways and tissue architecture

Teresa Attenborough[1,2]*, Kate A. Rawlinson[1,3]*‡, Carmen L. Diaz Soria[1], Kirsty Ambridge[1], Geetha Sankaranarayanan[1], Jennie Graham[1], James A. Cotton[1,4], Stephen R. Doyle[1], Gabriel Rinaldi[1,5], Matthew Berriman[1,2‡] (*These authors contributed equally, ‡ corresponding authors)

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, UK
2. Wellcome Centre for Integrative Parasitology, School of Infection and Immunity, College of Medical, Veterinary & Life Sciences, University of Glasgow, 120 University Place, Glasgow G12 8TA, UK
3. Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Massachusetts, USA.
4. Wellcome Centre for Integrative Parasitology, School of Biodiversity, One Health and Veterinary Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK
5. Department of Life Sciences, Aberystwyth University, Edward Llwyd Building, Penglais Campus, Aberystwyth SY23 3DA, UK

Abstract

Schistosomes are parasitic flatworms that cause the water-borne disease schistosomiasis, affecting millions of people worldwide. The miracidium larva of schistosomes represents the first post-embryonic stage of development and is critical to transmission. After hatching, a miracidium infects a freshwater snail and transforms into a mother sporocyst, where its stem cells generate daughter sporocysts that give rise to many human-infective cercariae larvae. To understand this important life cycle stage at the cellular and molecular levels, we have used single-cell RNA sequencing, *in situ* hybridisation and image analysis to create a whole-body cell atlas of the miracidium larva of *Schistosoma mansoni*. Our atlas shows that each miracidium is composed of ~365 cells and 19 transcriptionally distinct cell types. We show that 93% of miracidium cells are somatic (57% neural, 19% muscle, 13% epidermal (tegument), 2% parenchyma, 2% protonephridia), and the remaining 7% are stem cells. Cellular diversity within tissue types is revealed, and is highest in neurons. There are two stem cell populations, and they show different activation and potency states. Trajectory analysis indicates that one stem cell population is the origin of the tegument lineage and the

other likely contains pluripotent cells. Furthermore, each of these stem populations is transcriptionally distinct based on sex-linked gene expression in male and female larvae. Through single cell transcriptomics we identified every cell in the whole organism revealing the organisation of the miracidium, and present here the single cell atlas. Because of its simplicity, the miracidium provides a tractable system to understand the development of *Schistosoma mansoni* as it increases in complexity around the life cycle.

Introduction

The parasitic blood fluke *Schistosoma mansoni* is a causative agent of the neglected tropical disease schistosomiasis, which is responsible for a chronic disease burden equivalent to 1.0–2.6 million Disability Adjusted Life Years and over 11,000 human deaths annually (WHO, 2023; IHME, 2019). The life cycle of *S. mansoni* is complex, involving parasitic stages in two hosts (a snail and a mammal), with short-lived, freshwater infectious 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 in the host and continues over the ensuing six days (Ashton *et al*., 2001). During this time, the egg passes through the host vascular endothelium into the intestine and is passed in faeces into the outside environment (Costain *et al.*, 2018). If the egg lands in freshwater, a miracidium larva hatches out, swims and infects a snail host. Once inside the snail, the miracidium transforms into the mother sporocyst, and its stem cells (historically known as ‘germinal cells’) divide and differentiate to produce many daughter sporocysts that migrate within the snail. In the daughter sporocysts, a second round of embryogenesis generates the cercariae larvae by asexual reproduction from a stem cell. The cercariae emerge from the snail, swim towards and infect water-exposed mammals including humans, penetrating the skin. At the point of infection, the cercariae shed their tails, and once inside the mammal the tegument is remodelled during metamorphosis to the schistosomulum stage. The schistosomulum then moves into the blood vessels and, over the following five weeks, grows into a juvenile stage that migrates from the lungs to the liver and to the hepatic portal vein where sexually dimorphic male and female worms pair, mature and mate. Thus, during the development from egg to adult, the body axes are established three times, once during the development of the miracidia, once during the development of the daughter sporocyst and finally during the development of the cercariae. Additionally, there are two

rounds of embryogenesis, first from fertilised zygote to the mature miracidium, and second, the neo-embryogenesis from a single stem cell in the daughter sporocyst to a mature cercaria.

Enabled by the advent of single-cell RNA sequencing (scRNA-seq), molecular classification of cell types has been initiated for *Schistosoma mansoni* (Wang *et al*., 2018; Tarashansky *et al*., 2019; Li *et al*., 2020; Diaz *et al*., 2020; Wendt *et al*., 2020, Diaz & Attenborough *et al*., 2023). These 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 amongst stem cells that drive schistosome development and reproduction, classified a variety of differentiated cell types across the life cycle and revealed several key genes and their native cell types that are essential for parasite survival and propagation (Sarfati *et al*., 2021). Single-cell RNA-seq studies for the whole animal have only been carried out on intra-molluscan and intra-mammalian stages to date: the mother sporocyst (Diaz & Attenborough *et al*., 2023), the schistosomulum (Diaz *et al*., 2020), juvenile (3.5 weeks post-infection of mouse, Li *et al*., 2021) and adult (Wendt *et al*., 2020). Single cell studies on stem cells, however, have only 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).

An obvious gap in the coverage of the life cycle is the lack of single cell data for the earliest developmental time point currently accessible – the miracidium larva; earlier embryonic stages are inaccessible due to the impenetrable egg capsule. The miracidium is a short-lived, non-feeding stage that relies on internal stores of energy to sustain swimming, and host location and infection (Maldonado and Matienzo, 1948; Chernin, 1968, Pan, 1980). The role of the miracidium and its differentiated somatic cells is to carry stem cells to a favourable environment inside the snail where they can proliferate to generate daughter sporocysts, and subsequently cercariae. If the miracidium fails to infect a snail within 12 hours it dies.
At the ultrastructural level, a rich set of cell types have been previously identified in the miracidium and classified into eight major systems: musculature, nervous, penetration gland cells, excretory, interstitial, stem/germinal cells, epithelial (ciliary plates and epidermal ridges) and terebratorium (an apical modified epidermal plate with sensory structures and gland openings) (Pan, 1980).

To characterise this critical transmission stage, we have generated a single cell transcriptome atlas for the miracidium. Our atlas provides molecular definitions for all described cell types (Pan, 1980) and indicates the functions of different cell types within tissues. Multiplexed *in situ* validation of cell type marker genes enabled the cell types to be spatially resolved within the larva. Furthermore, the nuclear segmentation of the cells expressing tissue-specific genes has revealed the relative contribution of each tissue type to the animal and each cell in the larva has been assigned to a tissue. A key finding is the identification of two stem cell populations in a miracidium, and advances in our understanding of the genome (Buddenborg *et al*., 2021) has enabled us to determine that these two stem populations are transcriptionally distinct between male and female larvae due to sex-linked gene expression. To predict the fate of these two stem populations we carried out a trajectory analysis, combining scRNA-seq data from miracidia and mother sporocysts, and this indicated that one stem cell population likely contains the pluripotent cells that will develop into the cercariae, while the other population gives rise to the tegument lineage. Similarities in gene expression with the adult suggest that there is a tegument developmental program that is redeployed at multiple stages of the complex life cycle.

Results

Single-cell RNA-seq of 20,478 miracidia cells reveals 19 cell types.

There are ~365 nuclei in a miracidium (median = 365, larva #1: 342, larva #2: 365, larva #3: 377). However, some cell types, such as the apical gland and epithelial/tegument cells, are multinucleate; and others, such as the ciliary plates may be anucleate (Pan, 1980, but cf Meulemen *et al*., 1978). Therefore, the number of nuclei provides an approximation for the actual number of cells (Figure 1A).

We performed scRNA-seq on a pool of ~20,000 mixed-sex miracidia, collected within four hours of hatching from eggs (Figure 1B). Following dissociation, four samples of the cell suspensions were collected; two were enriched for live cells using fluorescence-activated cell sorting (FACS), and two were left unsorted to capture the maximum cellular diversity for the atlas (Figure 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 cells (Figure 1A), the number of quality-controlled cells theoretically represents >56-fold 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 (Figure 1C & D and Supplementary Table 2 and 3). Manually curated lists of previously defined cell-type-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 generated in this analysis. The expression of genes from both these lists in the miracidia single-cell data was then examined to identify the cell types that each Seurat cluster represented. Based on these markers, the 19 clusters were resolved into the following cell populations: two muscle-like (3,934 cells), five resembling neurons (3,255 cells), one ciliary plate (40 cells), one tegumental (366 cells), one protonephridial (539 cells), two parenchymal (1698 cells) and seven clusters resembling stem cells (10,686 cells) (Figure 1C).

Orthogonal body wall muscle fibres are transcriptionally distinct

The two muscle clusters were identified based on the expression of previously described muscle-specific genes; *paramyosin* (Smp_085540), *troponin t* (Smp_179810) and *titin* (Smp_126240)(Wendt *et al*., 2020; Diaz *et al*., 2020; Diaz & Attenborough *et al*., 2023), as well as other differentially expressed markers; e.g. a putative collagen alpha chain (Smp_159600)(Figure 2A). A crp1/csrp1/crip1 homolog (Smp_087250), encoding a conserved transcriptional regulator of muscle (Tarashansky *et al*., 2021) was also expressed in both muscle clusters. High expression of *paramyosin* (*PRM*; Smp_085540) was seen in muscle cell clusters 1 and 2 (Figure 2A). *In situ* hybridisation (ISH) (Figure 2B & C; Supplementary video 1) and image analysis (Supplementary video 2) identified 74 nuclei surrounded by *PRM* transcripts (i.e. ~19% of cells in a miracidium are muscle cells). Counterstaining with phalloidin (to label F-actin filaments) revealed an orthogonal grid of circular and longitudinal body wall muscles and from this we identified which PRM^+ nuclei belonged to each muscle subtype. The nuclei of the circular muscles formed two distinct bilaterally symmetrical lines that ran peripherally from pole to pole of the larva (~28 nuclei in total). Thirty-three more PRM^+ nuclei sat regularly spaced between the circular muscle nuclei and corresponded to longitudinal muscles. Another 13 PRM^+ nuclei formed a unilateral

cluster adjacent to the apical gland cell (identifiable by its four nuclei; Pan, 1980) (Figure 2Cii; Supplementary Video 1 and 2).

Muscle cluster 1 (2,407 cells) was distinguished from muscle cluster 2 by markedly higher expression of a Kunitz-type protease inhibitor (Smp_052230; Figure 2A), and ISH showed Smp_052230 transcripts along the two peripheral lines of circular muscle nuclei (Figure 2D). Subpopulations of muscle cluster 1 also expressed orthologs of *Wnt-11-1* (Smp_156540) and *Notum* (Smp_015820), known anterior-posterior body axis patterning genes, where Notum acts as an inhibitor of Wnt signalling (Peterson and Reddien, 2009; Kakugawa *et al.*, 2015)(Figure 2A). *Notum* (Smp_015820) was highly expressed in the cluster of 13 muscle nuclei adjacent to the apical gland cell (Figure 2E). *Wnt-11-1* (Smp_156540) was expressed at the posterior pole, in seven of the Kunitz-type protease inhibitor-expressing circular muscle cells (Figure 2D; Supplementary video 3). *Notum* was also expressed, albeit weakly, in some nuclei of circular muscles, including those expressing *Wnt-11-1* (Supplementary Video 4).

Muscle cluster 2 (1,355 cells) was distinguished by higher expression of genes encoding a PDZ domain protein (Smp_311170) and *Calcium-transporting ATPase* (Smp_136710), as well as the unique expression of a putative calcium-activated potassium channel (Smp_156150), and the transcription factor *MyoD* (Smp_167400)(Figure 2A); ISH of the latter two genes showed expression only in the longitudinal body wall muscles (Figure 2F; Supplementary video 5). Orthologs of other axial patterning genes, *Netrin receptor UNC5* (Smp_172470), *Wnt5* (Smp_145140), and a putative wnt inhibitor frizzled, frzb2 (Smp_062560)(Witchley *et al*., 2013), showed distinct expression in muscle 2 cells.

The higher expression of muscle 1 markers in circular body wall muscles (BWM) and muscle 2 genes in the longitudinal BWM revealed distinct transcriptomic signatures for these two types of myocyte that make up the orthogonal body wall musculature. In fact, differential expression analysis showed that *Wnt-11-1* (circular) and *MyoD* (longitudinal) were among the most differentially expressed genes 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)(Figure 3A). *Complexin* was expressed in and around >209 nuclei, indicating that 57% of cells in the miracidium were neurons; out of which 129 *cpx*$^{+}$ nuclei formed the nuclear rind of the brain (or neural mass/ring), and the remaining 80 were situated peripherally, either anterior or posterior to the brain (Figure 3B). Sensory organelles were identified for peripheral neural cell types.

Cells from neuron clusters 1, 2, and 3 had nuclei in the nuclear rind of the brain. Neuron cluster 1 was large (1,470 cells), connected to stem F and expressed neuron-specific *complexin*, the neuroendocrine precursor *7B2* (Smp_073270) and an uncharacterised gene (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), fifteen subclusters could be resolved (Supplementary Figure 1; Supplementary Table 5). Three of the subclusters expressed the following neuropeptide precursor genes amongst their top five markers and expressed in the nuclear rind of the brain (Supplementary Figure 2): *PWamide* (Smp_340860)(Wang *et al*., 2016) in cluster 14, *Neuropeptide F* (Smp_088360) in cluster 8, and *Neuropeptide Y* (Smp_159950) in cluster 13. Another subcluster was composed of putative photoreceptor cells (cluster 7 - Supplementary Figure 1, Supplementary Table 5) because their top-ten markers include two opsins (Smp_180350 and Smp_332250) – identified here as a rhabdomeric opsin and a peropsin, respectively (Supplementary Figure 3) – and other conserved phototransduction cascade genes, such as a 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase (Smp_245900) and a beta-arrestin (Smp_152280) (Supplementary Figure 1, Supplementary Table 5).

Two further rhabdomeric opsins (Smp_180030 and Smp_104210; Supplementary Figure 3) were expressed in the Neuron 2 cluster, but across the cluster of 282 cells, the most expressed marker was the uncharacterised gene Smp_201600. ISH revealed three strongly Smp_201600^{+} cells, with nuclei in a median posterior location of the brain, and three cells with weaker expression in a more anterior position (Figure 3C, Supplementary video 6). Other top markers include *7B2*, a guanylate kinase (Smp_137940) and a putative homeobox protein (Smp_126560).

Neuron 3 was a small cluster (47 cells) expressing *complexin* and *7B2*. The uncharacterised gene, Smp_071050, was unique to and highly expressed throughout this population (Figure 3A). ISH showed expression in ~6 cells with projections that extended into the brain and to

the periphery, three of which sat adjacent to three neuron 2 cells. There was also weak expression in the four nuclei of the apical gland cell (Figure 3Ci, Supplementary Video 6). STRING analysis of neuron 3 markers produced two networks with functional enrichment; these were associated with 'synapses' and 'calcium' (Supplementary Figure 5). In neuron clusters 1, 2 and 3, there was an enrichment for genes annotated with Gene Ontology (GO) terms related to 'neuropeptide signalling pathway' and 'neurotransmitter transport' (Supplementary Table 6). There were, therefore, at least 3, and up to 15 more, transcriptionally distinct cell types in the brain mass.

Cells from neuron clusters 4 and 5 had nuclei peripheral to the brain. Neuron 4 formed a large (1,140 cells), discrete cluster. The top marker was an uncharacterised gene, Smp_319480 (Figure 3A), that, based on protein-structure prediction using I-TASSER (Yang & Zhang, 2015), was similar to human mTORC1 (PDB entry 5H64; TM-score 0.85). *In situ* expression revealed ~26 Smp_319480^{+} cells; four at the anterior end, with projections leading to the body wall in the first tier of ciliary plates, and ~22 sat 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 (Figure 3D, Supplementary Video 9 and 10). Twenty-three multi-ciliated nerve endings formed a girdle around the larva between the second and third tier of ciliary plates (Supplementary Figure 2C), and Smp_319480 transcripts were expressed at the base of these cilia (Supplementary Figure 2D); some transcripts also reached the nuclei of circular body wall muscles (Supplementary Figure 2E), suggesting these cells may function as sensory-motor neurons. Two genes, identified as top markers and co-expressed in many neuron 4 cells, encode likely orthologues of *Polycystin 1* (Smp_165230) and *Polycystin 2* (Smp_165660)(Supplementary Figure 4A), with the latter prediction strongly supported by a Foldseek search (van Kempen *et al*., 2023) against the protein database (PDB). Polycystins 1 and 2 form a heterodimeric ion channel required for oscillating calcium concentrations within vertebrate cilia (Kim *et al*., 2016). A Sox transcription factor, Smp_301790, was also expressed in neuron 4 and co-expressed with *Polycystin 1* and *2* (Supplementary Figure 4B). Other top markers included a putative potassium channel (Smp_141570), a cAMP-dependent protein kinase (Smp_079010), a GTP-binding protein (Smp_045990) and an ortholog of human *fibrocystin-L* (Smp_303980). The GO terms 'cilium', 'regulation of apoptotic process' and 'potassium ion transmembrane transport' were significantly enriched, and

STRING analysis showed networks enriched in ‘plasma membrane-bound cell projection’ (Supplementary Table 6. Supplementary Figure 5).

Neuron 5 cluster comprised 276 cells. An EF-hand domain-containing protein, Smp_343710, was highly expressed in 55% of these cells. ISH showed expression in 10-20 cells whose nuclei sat outside of, and anterior to, the brain, in and around the paired lateral glands and their secretory ducts (Supplementary Video 11). There was also expression in a pair of bilaterally symmetrical bulbous protrusions (called lateral sensory papilla and thought to be depth sensors, Pan, 1980)(Figure 3F). STRING analysis revealed networks associated with ‘synapses’, ‘ion channel binding’ and ‘cation transmembrane transporter activity’ (Supplementary Figure 5).

We expected to find a discrete cluster(s) for the penetration glands within the neural clusters (Steger *et al*., 2022); there are three gland cells per larva - one apical and two lateral (Pan, 1980). While the top marker gene for neuron 3 was expressed in the four nuclei of the apical gland (Figure 3Ci), and there was expression of a neuron 5 marker in the lateral glands (Figure 3Fi), these genes were expressed in many other cells as well. This suggests that gland cells are either sub-clusters of neurons 3 and 5 or they were not captured in the 10X GEMs. To characterise the transcriptomes of the gland cells, we carried out plate-based single cell RNA-seq (New England BioLabs) on 37 manually selected gland cells (Supplementary Table 7). The biological process GO terms enriched in the most abundant genes (top 200 genes by median transcripts per million, Supplementary table 7) included translation, protein folding and ATP synthesis coupled electron transport (Supplementary Table 8). The gland cells showed co-expression of neural markers (*Complexin* and *7B2*) and muscle markers *Myosin light chain 1* (Smp_045200), *Troponin t* (Smp_179810) and *Actin-2* (Smp_307020) (Supplementary Figure 6). To identify genes that were specific to gland cells, we searched for differentially expressed genes in the gland cells compared to the other cells in the plate-based scRNA-seq data (i.e. ciliary plates and unknown cells) (Supplementary Table 9 and 10), and then looked at their expression in the 10X-based scRNA-seq data. A lack of, or minimal, expression of these markers in the 10X-based scRNA-seq data, would support the hypothesis that they were specific to the gland cells. Five venom antigen-like (VAL) genes (belonging to the SCP/TAPS or CAP) superfamily) were expressed in the gland cells; Smp_120670 (VAL 5) and Smp_070250 (VAL 15) in over 60% of the gland cells, and Smp_176180 (VAL 9), Smp_002630 (VAL 2) and Smp_331830 (SCP protein) in less than 50%. Many gland cells

expressed multiple VAL genes (Supplementary Figure 6). In the 10X-based scRNA-seq data, very few cells expressed these VAL genes (1-5 cells for each gene, distributed between neurons 1 and 5, muscle 1, parenchyma 2, and stem F and G clusters). The VAL genes expressed in the gland cells are uniquely expressed during the egg, miracidia and sporocyst stages (Yoshino *et al*., 2014; Lu *et al*., 2018; Buddenborg *et al*., 2023), and it has been reported that VAL 9 (Smp_176180) induces the differential expression of an extracellular matrix remodelling gene in the snail host (Yoshino *et al*., 2014). Other differentially expressed genes in the gland cells included Smp_303400 (an HSP70 homolog), Smp_179250 (*α-galactosidase/α-N-acetylgalactominidase*), Smp_317530 (uncharacterised protein), and two putative major egg antigens Smp_185680 and Smp_303690 (Supplementary Figure 6). Eight putative secreted proteins previously identified in miracidia and *in vitro* transformed miracidia (Wang *et al*., 2016; Wu *et al*., 2006) were expressed in the glands cells (and not in other cell types based on the 10X data), and five of these showed relatively high expression; Smp_302170 (*HSP70* homolog), Smp_303400 (*HSP70* homolog), Smp_049230 (*HSP20*), Smp_304250 (*purine nucleoside phosphorylase*), and Smp_302180 (putative heat shock protein 70)(Supplementary Figure 6). Typical of venom glands in other animals, the transcriptomes of the miracidia gland cells are enriched in genes involved in protein translation, stabilisation and stress response to cope with mass protein production (Zancolli *et al*., 2021).

Ciliary plate gene marker identifies six submuscular cell bodies

A small cluster of 40 cells (almost exclusively from the unsorted samples) was identified near the tegument cluster (Figure 1C). The top markers were Smp_096390, a 16 kDa calcium-binding protein (CaBP) and *p25 alpha* (Smp_097490, a marker for ciliated neurons in adult worms, Wendt *et al*., 2020) (Figure 3A). *In situ* hybridisation revealed CaBP and *p25 alpha* expression in the ciliary plates; epithelial cells with thousands of motile cilia used for swimming (Figure 3G and Supplementary Figure 2). Both genes were also expressed in six previously undescribed cells that sent a projection out to the exterior (between the first and second tier of ciliary plates as seen with phalloidin staining, Figure 3G) and whose nuclei sat below the body wall musculature (Supplementary Videos 7 & 8; Supplementary Figure 2F). Given the association of *p25 alpha* with ciliated neurons in adults (Wendt *et al.*, 2020), expression in the sub-muscular cells suggested they project cilia into the external

environment and may be sensory cells. The expression of CaBP and *p25* in both the ciliary plates and submuscular cells suggests that these anatomical structures are functionally linked. GO terms for this cluster were related to 'mitochondrial transmembrane transport' and 'microtubule-based process' (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 5). Further plate-based scRNA-seq of 21 individual ciliary plates confirmed the high expression of CaBP (Smp_096390) and *p25 alpha* (Smp_097490), with more than two-fold higher expression of CaBP than any other gene by median TPM (transcripts per million) (Supplementary Table 11). Although the ciliary plates expressed *p25* (a ciliated neuron marker in the adult, Wendt *et al.*, 2020), they did not express many of the typical neural markers, e.g. *complexin* and *7B2* (Supplementary Figure 6).

Tegumental cells expressed *MEG6* and are located in the posterior two-thirds of the larva.

We identified a cluster of 366 cells as tegumental based on the expression of markers for tegument in the adult worm (Wendt *et al.*, 2020), e.g. genes encoding *dynein light chain* (Smp_302860, Smp_200180), a DUF3421 domain-containing protein (Smp_331910) and micro exon gene *MEG6* (Smp_163710). Two other dynein light chain genes were top markers (Smp_201060, Smp_312630), and both a tetraspanin (Smp_303290) and *Gelsolin* (Smp_008660) were specific to this cluster. *Meg6* (Smp_163710) was expressed in 66% of the cells in this cluster; ISH showed 46 *Meg6*$^{+}$ cells, and the nuclei were in the posterior two-thirds of the larva. The nuclei sat below the body wall muscle, and cytoplasmic protrusions reached between muscle filaments and formed the epidermal ridges between the ciliary plates (Figures 4E & F). These cells correspond to the epidermal ridge cells described by Pan (1980). A second tegument marker, a dynein light chain (Smp_200180), showed overlapping expression with *Meg6* (Supplementary Video 15). Enriched GO terms for tegument genes included 'microtubule-based process' and 'protein folding' (Supplementary Table 6). STRING analysis revealed a network of 25 genes associated with protein localisation to the ER, folding, export and post-translational protein targeting to the membrane (Supplementary Figure 7).

Spatial localisation of top marker transcripts identifies a simple protonephridial system

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). Eighty percent of the cells expressed the protonephridia marker Smp_335600 (ShKT-domain containing protein) (Sarfati *et al.*, 2021) (Figure 4A). Smp_335600 was expressed around six nuclei in a bilaterally symmetrical pattern (Figure 4Bi) (Supplementary Video 12). On each side, Smp_335600 transcripts were distributed along the s-shaped path of an excretory tubule, passing its nucleus, and connecting the anterior and posterior flame cells (Figure 4Ci). Within the flame cells, expression was detected around the nuclei rather than the barrel (Figure 4Cii). The four ciliated flame cells and two excretory ducts comprise the excretory system of the miracidium. Accordingly, STRING (Supplementary Figure 7) and GO analyses (Supplementary Table 6) of this cluster confirmed enrichment for terms including 'cilium organisation', 'axoneme assembly', 'cytoskeleton organisation' and 'microtubule-based process'.

Two distinct parenchyma clusters predicted in the miracidium

We identified two clusters of cells (1,093 cells from parenchyma 1 and 605 cells from parenchyma 2) that were transcriptionally similar to the parenchymal cells found in later developmental stages; e.g. they express hypothetical protein (Smp_063330), *cathepsin B-like cysteine proteinase* (Smp_141610) and *serpin* (Smp_090080) (Diaz *et al.*, 2020; Wendt *et al.*, 2020; Diaz & Attenborough *et al*., 2023). All top marker genes for parenchyma 1 were also expressed in parenchyma 2 but at lower levels, whereas the DUF2955 domain-containing protein (Smp_191970) was uniquely expressed in parenchyma 2 (Figure 4A). A gene encoding a Zinc transporter Slc39a7 (Smp_318890) was a top marker for both clusters, with expression confirmed *in situ* within 7–9 cells; two of which were located anteriorly on either side of the brain and 5-7 located posterior to the brain (Figures 4B & F). These parenchyma cells have long cytoplasmic protrusions that reach between all the other cells, filling the intercellular space (Figure 4Fii-iii; Supplementary Videos 13 & 14). Pan (1980) called these the interstitial cells and, observing abundant glycogen particles and lipid droplets, proposed they function as sources of energy for the non-feeding larva and as nurse cells for the stem cells during the first few days post-infection of the snail. 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 could only identify parenchymal cells in general and could not identify the two clusters *in situ*. STRING analysis for the parenchyma 1 showed enrichment in networks for 'carbohydrate metabolic process' and 'hydrolase', and for parenchyma 2: 'phagosome', 'mTOR signalling pathway', 'mitochondrial inner membrane' and 'proton-transporting two sector atpase complex' (Supplementary Figure 8). Analysis of GO annotations was able to distinguish between the two parenchyma sub-populations, with parenchyma 1 genes enriched for 'extracellular region' and 'endopeptidase inhibitor activity', and parenchyma 2 genes enriched for processes that are 'dynein complex' and 'microtubule-based process' (Supplementary Table 6; Supplementary Figure 9).

Two major stem cell populations in the miracidium that clustered by sex

We identified stem cells based on the expression of known 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)(Wang *et al*., 2013; 2018)(Figure 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)(Figure 5A, Supplementary Figure 10); 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 (Figure 5B).

Working with the latest version of *S. mansoni* genome (v10), where the sex chromosomes Z and W have been resolved (Buddenborg *et al*., 2021), we were able to determine that the Delta/Phi-like and Kappa-like populations are each composed of two well-separated clusters that can be distinguished based on the expression of sex-linked genes (Figure 5A; Supplementary Figure 11). *Schistosoma mansoni* has 7 pairs of autosomes and one pair of sex chromosomes; males are homogametic (ZZ) and females heterogametic (ZW) (Buddenborg

et al., 2021). The sex chromosomes are composed of sex-specific regions that are unique to each chromosome flanked by pseudoautosomal regions that are common to both Z and W. To unveil the genes contributing to the sex-driven clustering we removed in turn different sex chromosome regions from the analysis. When all Z-specific region (ZSR) and W-specific region (WSR) genes were removed simultaneously, the two Delta/Phi-like and two Kappa-like clusters each collapsed into single stem cell clusters, giving five stem cell clusters in total (i.e. Delta/Phi, Kappa, E, F and G; Figure 5B). Removing from the dataset genes located solely on the WSR, the ZSR, or just the gametologs (i.e. genes with homologous copies on the WSR and ZSR) did not impact the overall UMAP clustering, i.e. in all three scenarios the stem cell clusters remained separated by sex (Supplementary Figure 11). ZSR or WSR expression is therefore sufficient on their own to split each of these stem clusters in two (Supplementary Figure 11).

The unexpected subclustering of Delta/Phi and Kappa stem cells by sex could be due to an incomplete or a lack of dosage compensation in the ZSR genes (Supplementary Figure 11A&B). In many animals, dosage compensation equalises gene expression across the sex chromosomes that are unevenly distributed between the sexes. We classified all cells based on the expression of WSR genes (i.e. only found in female individuals), so each cell was designated as W^+ (likely female) or W^- (likely male) (Supplementary Figure 12A&B). The difference in ZSR and WSR gene expression was most pronounced in the Delta/Phi-like and Kappa-like cells, shown by the separation of W^+ and W^- cells (Supplementary Figure 12B). However, this analysis revealed other tissue-specific sex differences in ZSR gene expression e.g. in the protonephridia, which was evident through segregation of W^+ and W^- cells within that cluster. We found that W^+ and W^- cells were well mixed in the majority of the remaining clusters (Supplementary Figure 12B). Male cells have two copies of the ZSR (ZZ) and produce higher coverage across genes in this region compared to female cells, which have one ZSR and one WSR (ZW). These findings suggest that sex-linked gene dosage is not fully compensated in the miracidium, and there may be tissue-specific variation.

To identify the stem cells *in situ*, we investigated the expression of the pan-stem marker *ago2-1*. *Ago2-1*$^+$ cells were distributed lateral and posterior to the brain (Figure 5Ci) and

there was variation in the number between larvae, ranging from 20-29 (n = 5, median = 23, mean = 23.4)(i.e. ~7% of cells that make up the larva are stem cells). Cells in the Delta/Phi-like cluster expressed a canonical *p53 tumour suppressor family gene* (*p53-1*, Smp_139530) as a top marker; *in situ* expression showed 12–14 *p53-1*$^{+}$ cells (n=3)(Figure 5D). Cells in the Kappa-like cluster expressed *Uridine phosphorylase A* (*UPPA*, Smp_308145) as a top marker; *in situ* expression showed 8–9 *UPPA*$^{+}$ cells, and they were located amongst the *p53-1*$^{+}$ cells (n = 3)(Figure 5D). We were not able to detect cells from the stem E, F and G clusters in the miracidium using ISH as there were no predicted marker genes that were unique to these clusters. Although *Phosphomevalonate kinase* (*pmk*) (Smp_072460) was a predicted marker for the stem E cluster (Figure 5A), when multiplexed as *p53/UPPA/pmk*, no *pmk*$^{+}$-only cells could be found (Supplementary Video 16). Likewise, multiplexing *ago2-1/ p53/ UPPA* revealed no *ago2-1*$^{+}$-only cells (Supplementary Video 17).
Overall, we could spatially validate two stem cell populations within the miracidia: Delta/Phi and Kappa. Cells in stem E, F and G *in silico* clusters might be stressed/damaged/dying cells or cells in transcriptionally transitional states.

A lack of EdU incorporation into miracidia cells (Supplementary Figure 13) indicated that DNA replication does not take place in the stem or somatic cells during this free-swimming stage. 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 the Delta/Phi and Kappa stem clusters were ribosomal proteins (Figure 5A). STRING analysis showed that most of the top 100 markers of each cluster both formed large single networks enriched for genes involved in translation, ribosome biogenesis and gene expression, and this was particularly striking in the Delta/Phi cluster (Supplementary Figure 14). Amongst the top genes differentially expressed between Kappa and Delta/Phi (Supplementary Figure 15A; Supplementary Table 12) were six transcription factors up-regulated in Delta/Phi: an ortholog of *lin-39* (Smp_247590), *tiptop* (Smp_180690), *hes-related* (Smp_024860), *p53-1* (Smp_139530), *Zfp-1* (Smp_145470) and a putative *homothorax homeobox protein/ six-3-1* (Smp_147790). The simultaneous co-expression of

multiple transcription factors, indicates a multipotent progenitor stage (Kumar *et al*., 2019). Enriched in, 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 a very long chain fatty acid elongase (Smp_051810). GO analysis of differentially expressed genes (Supplementary Table 13) revealed that the most significantly upregulated genes in Delta/Phi were related to the structural component of the ribosome and translation; these are the first markers of a pre-activation stage in stem cells (Shin *et al*., 2015). Whereas GO analysis showed that up-regulated genes in Kappa are involved in transcription and metabolism (Supplementary Figure 15B), which indicated quiescent rather than primed and activated stem cells (Llorens-Babadilla *et al*., 2015).

One miracidia stem population contributes to somatic tissues and the other to the germline.

The Delta/Phi stem population simultaneously co-expressed three transcription factors that are known to be specific to the flatworm tegument/epidermal lineage, i.e *p53-1, zfp-1, six3-1*(Cheng *et al*., 2018; Collins *et al*., 2016; Wendt *et al*., 2018; 2022). This suggests that the Delta/Phi stem population may have been the origin of the miracidia tegument cells and may be the source of new tegument cells in the sporocyst stages. To explore the fates of the two miracidia stem populations, we investigated the dynamics of their gene expression in stem and tegument cells across the transition from miracidia to sporocyst. To do this, we combined single-cell data from these developmental stages and cell types (this study and Diaz & Attenborough *et al*., 2023) and used RNA velocity analysis (Bergen *et al*., 2020) to estimate relative induction/repression and timing of gene expression in this dataset.

A UMAP plot of the combined clustering shows that the miracidia tegument cells form a bridge connecting the miracidia stem Delta/Phi with the sporocyst tegument cells (Figure 6A) (batch-corrected samples showed similar results, Supplementary Figure 16). The velocity field and latent time representations show the progression of some miracidia Delta/Phi stem cells towards the miracidia tegument cells and a high speed of differentiation was inferred from the miracidia tegument cells to the sporocyst tegument (Figure 6B & C). This suggests

that the miracidia tegument cluster may contain cells that range from post-mitotic stem cell progeny to tegument progenitors to terminally differentiated tegumental cells.

Investigations into tegument development in adult schistosomes has revealed key regulators and genes associated with different development states (Collins *et al*., 2016; Wendt *et al*., 2018; Wendt *et al*., 2022). We examined the dynamics of these genes in the combined miracidia/mother sporocyst dataset using RNA velocity and found that the velocities of these genes show different temporal dynamics. *P53-1* (Smp_139530) was induced in the Delta/Phi stem cells but repressed in some miracidia tegument cells and expression decreased along the bridge formed by the miracidia tegument cells (Figure 6D). The flatworm-specific transcription factor *Zfp-1* (Smp_145470) showed a similar expression pattern to *p53-1*, except that the predominance of spliced copies (Figure 6E) suggested that transcription happened earlier, during embryogenesis. A second zinc finger protein, *Zfp-1-1* (Smp_049580) was expressed in tegumental cells more than stem cells, which suggested a role in stem cell progeny and tegument progenitors (Supplementary Figure 17). In adults, tegument cells are renewed continuously by a population of *tsp2+* (Smp_335630) progenitor cells that fuse with the tegument (Wendt *et al*., 2018). In the miracidia, *tsp2* was expressed in the Delta/Phi stem cells as well as tegument cells (Supplementary Figure 17), suggesting earlier expression than in the adult, or that Delta/Phi are progenitor cells. Several markers of mature tegument cells in the adult were expressed in the miracidia and sporocyst tegument cells but not in the stem cells, e.g. *annexin* (Smp_077720), *calpain* (Smp_241190) and *endophilin B* (Smp_003230) (Supplementary Figure 17). *Annexin* is specific to the tegument cells, and the phase plot showed the accumulation of mature transcripts in the sporocyst tegument, while there was evidence of recent transcription in the miracidia tegument (Supplementary Figure 17). A putative ortholog of a planaria transcription factor *six-3-1* (Smp_147790), a marker for early epidermal progenitors (Cheng *et al*., 2018) was expressed in the Delta/Phi stem cells and miracidia tegument (Supplementary Figure 17). All of these genes showed little expression in the Kappa stem population (Supplementary Figure 17). The expression of these known tegument/epidermal development and lineage-specific genes in the Delta/Phi stem cluster, and the indication that *zfp-1* transcription occurred pre-hatching, suggested that this stem population was the origin of the miracidia tegument cells and is the

source of the tegumental lineage. This, together with the observation that more than 80% of protein-coding genes in the genome are detected in these miracidia data, suggests that similar molecular programs for the development of tissues in the miracidia and mother sporocyst may be redeployed multiple times across the life cycle.

As well as allowing us to interrogate the dynamics of known marker genes (Supplementary Figure 17), the RNA velocity analysis also revealed genes previously unknown to have dynamic behaviours across these stages; these included genes up-regulated from miracidia tegument to sporocyst tegument, e.g. *epidermal growth factor receptor* (Smp_035260), genes down-regulated from miracidia tegument to sporocyst tegument, e.g. *galectin domain* (Smp_140590), and dynamic behaviour across stem cells and tegument, e.g. *unc44/ankyrin 2,3* (Smp_145700)(Supplementary Figure 17, Supplementary Table 14).

In the Kappa stem population, there was transcription and expression of genes detected in germline stem cells (GSCs) and germ cells in the intra-mammalian stages of the life cycle (Wang *et al.*, 2018; Li *et al.*, 2021). A second P53 homolog, *p53-2* (Smp_136160), was highly expressed and actively transcribed in Kappa cells (Figure 6F). In adults, this gene is enriched in reproductive organs and has a genotoxic stress response that is thought to defend the integrity of the genome by eliminating germ cells that have acquired mutations (Wendt *et al*., 2022). Also expressed solely in Kappa cells are the germline-specific regulatory genes *eled* (Smp_041540)(Figure 6G), *onecut1* (Smp_196950), *irx* (Smp_063520), *akkn* (Smp_131630), and *boule* (Smp_144860)(expressed in the male Kappa cells and stem E, Supplementary Figure 20)(Wang *et al.*, 2018; Li *et al*., 2021), as well as *klf* (Smp_172480) (Supplementary Figure 18), a homolog of KLF4, which is an important transcriptional regulator governing pluripotency (Takahashi & Yamanaka, 2006).

Together, these data suggest that the Delta/Phi population is the origin of the tegumental lineage and likely contributes to extra-embryonic somatic tissues during intra-molluscan development, while the Kappa population likely contains pluripotent cells that will contribute to the development of cercariae and the germline in intra-mammalian stages (Figure 7).

Discussion

Schistosomes are complex metazoan parasites, with adult worms of *Schistosoma mansoni* characterised into 68 transcriptionally distinct cell clusters (Wendt *et al*., 2020). Understanding the genes involved in the development and maintenance of these cell types is a promising approach for identifying novel therapeutic targets (Wendt *et al*., 2020). Here we have shown that the miracidium larva is composed of ~365 cells. Compared to later life cycle stages, where the cercariae is made up of ~1,600 cells, the schistosomulum ~900 (Diaz *et al*., 2020) and the adults tens of thousands of cells (unpublished data), the miracidium is a simple stage of the life cycle. We identified and spatially resolved 19 transcriptionally distinct cell types, revealing the relative contribution of each tissue type to the larva (Figure 8). The miracidium, therefore, with its diversity of differentiated cells present in relatively low numbers, makes a simple and tractable system with which to gain a fundamental understanding of the biology and spatial architecture of schistosome cells, and their transcriptomes.

Nineteen cell types may be a conservative characterisation, but at this resolution most cell types formed clearly defined clusters. There was sufficient sensitivity to detect cell types with as few as six cells per larva, for example, neurons 2 and 3, and the protonephridial system. Some rarer cell types are embedded in existing clusters, particularly in neuron 1, and a complete neural cell atlas and a study of the putative photoreceptor cells is in progress (Rawlinson *et al*., in prep). Some cell types were preferentially recovered, e.g. the stem cells; ISH revealed that they 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. The ciliary plates, for instance, were only found in sufficient numbers to be defined as a cluster in unsorted samples. 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 known cell types were not captured, i.e. the

penetration glands. With three gland cells per larva and a theoretical 56x coverage of each cell, we would have expected ~168 gland cells in the dataset. These cells were easily identified after dissociation by their distinct flask-shape but were not identifiable in these 10X data. Perhaps, because of their large size (20x50µm) and irregular shape, they were not incorporated into the 10X GEMs. Therefore, we used plate-based scRNA-seq for the penetration glands to complete the cell type transcriptome atlas for all known cell types. Together, we have transcriptomically characterised the anatomical systems described by Pan (1980), and we discuss key findings below.

The two muscle clusters in our dataset correspond to the circular and longitudinal muscles that comprise the orthogonal body wall muscles of the miracidium. 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). Posterior Wnt signalling and anterior Wnt inhibition are commonly observed features of anterior-posterior body patterning in animals (Peterson & Reddien, 2009). The expression of *Wnt-11-1* and a Wnt inhibitor, *Notum*, in the circular muscles at opposite poles of the miracidium suggested these muscles are involved in defining the anterior-posterior axis. 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 of which have been identified in the miracidium before. Many 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 stem cells during homeostasis in the adult (Wendt *et al*., 2020). As the body axes of the miracidium were defined during embryogenesis, the expression of these genes in the muscles of the larva suggest they will provide positional information to the stem cells of the mother sporocyst in the axial patterning of the daughter sporocysts.

The nervous system accounts for over 50% of cells that make up the larva and is the most transcriptionally heterogeneous tissue, as 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. This is evident in our scRNA-seq data by the presence of discrete, well-defined neural cell clusters. *In situ* validation of marker genes revealed many of the nerves identified ultrastructurally; for example, the location and shape of neuron 4 cells resemble that of the bipolar ganglion cells with multi-ciliated nerve endings described by Pan, 1980. Examination of their transcriptional profiles indicates their function; top markers include a putative *mTORC1*, two *Polycystin* genes, a *GTP-binding ras-like* and *fibrocystin.* Mammalian orthologs of these 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 1 and 2 expressed in ciliated neurons of the larva of the annelid *Platynereis dumerilli* are involved in mechanosensation, detection of water vibrations and mediate the startle response (Bezares-Calderón *et al.*, 2018). Furthermore, in planaria, SoxB1-2 regulates polycystin genes and together they function in ciliated cells that are involved in water flow sensation (Ross *et al*., 2018). The gene expression profile of neuron 4 cells in the miracidium suggests they could function as flow and/or vibration sensors and may be specified by a Sox transcription factor. These multiciliated receptors degenerate 12 hours after the miracidia transform into a mother sporocyst (Poteaux *et al*., 2023).

In situ validation of marker genes also revealed an undescribed cell type, namely the six submuscular cells that express the top marker genes for the ciliary plates (i.e. a 16KDa calcium binding protein and *p25 alpha*). The expression of these genes in both the ciliary plates and the submuscular cells suggests that these latter cells are transcriptionally similar and might play a role in the development and functioning of the ciliary plates. Their projection into the external environment suggests a sensory role, which, together with their location between the first and second tier of ciliary plates, could suggest a role in coordinating the movement of the motile cilia across the plates and tiers in response to sensory cues (i.e. similar in function to the ciliomotor pacemaker neurons in *Platynereis* larvae, Verasztó *et al*., 2017). The ciliary plates are a miracidium-specific cell type (it is the only life cycle stage that uses motile cilia for locomotion), and once inside the snail, these plates are shed within 2 – 48 hours (Meulement *et al*., 1978; Poteaux *et al*., 2023). As expected, they are transcriptomically typified by enrichment with dynein and high metabolic

activity. The 16KDa calcium binding protein (Smp_096390) is clearly important to the function of ciliary plates as it is highly expressed in data collected from both droplet- and plate-based scRNA-seq platforms. Accordingly, bulk RNAseq data shows this gene was previously detected in the late embryonic stage, miracidia and early mother sporocyst stage (Lu *et al*., 2021; Buddenborg *et al*., 2023).

Between all the ciliary plates sit the ridges of the tegumental cells. The ridges connect to the cell bodies beneath the body wall musculature, and their transcriptomes show a predicted network of genes involved in protein folding, export and translocation. This is consistent with ultrastructural studies that have observed abundant RER, ribosomes, Golgi apparatus, and complex carbohydrates in these cell bodies, suggesting they are active in producing exportable proteins (Meulement *et al*., 1978; Pan, 1980). Large numbers of membrane-bound vesicles were also observed and thought to become part of the new tegument membrane (syncytium) that rapidly forms around the mother sporocyst during the early stages of transformation, protecting the mother sporocyst from attack by the snails' amoebocytes (Pan, 1980). The RNA velocity analysis has revealed dynamic genes that may be crucial in the maturation of the tegument cells across this life cycle transition, and in turn, for evasion of host defences and parasite survival. This analysis also suggests a developmental trajectory for the tegument cells that starts in the miracidia Delta/Phi stem cells and follows their progeny to terminally differentiated tegument cells in the miracidium and five-day-old sporocyst. Many of the molecular regulators and markers of adult tegument biogenesis (Collins *et al*., 2016; Wendt *et al*., 2018; Wendt *et al*., 2022) are expressed in the miracidia and sporocyst stem and tegument cells, indicating that there is a conserved molecular program for the development of the syncytial tegument in the mother sporocyst and its maintenance in the adult worm.

The transcriptomes of the two stem populations in the miracidium suggest they are in different activation states. 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 transforms into the mother sporocyst inside the

snail. These two stem populations may also have different potency states; the co-expression of multiple transcription factors in Delta/Phi, including those specific to the tegument lineage (Wendt *et al*., 2018; 2022), indicates a multipotent progenitor stage. Whereas in the Kappa stem cells, the expression of genes that define the germline in the intra-mammalian stages (Wang *et al*., 2018) suggest pluripotency. Supporting this, the RNA velocity analysis shows that *p53-2*, the genome protection gene (Wendt *et al*., 2022), is most highly expressed, and being actively transcribed, in the Kappa population. The incorporation of miracidia single-cell data into existing stem cell scenarios on *Schistosoma mansoni* development (Wang *et al.*, 2013; Wang *et al*., 2018; Li *et al*., 2021) shows the continuum of the Kappa population from hatching through the intra-molluscan stages to the adult stage inside the mammalian host (Figure 7). The finding that *eled*, and many of the germline-specific regulatory program components, are expressed in some miracidia Kappa cells suggests that when a miracidium hatches, the cells that will contribute to the germline may already be segregated and the molecular regulatory program that differentiates somatic (Delta/Phi) and germ cell (Kappa) lineages is active.

An unexpected finding regarding the stem cells was the further clustering of the two populations by sex. This is the first identification of any sex differences in the miracidium. This degree of transcriptomic sexual difference was most evident in the Delta/Phi and Kappa stem cell clusters, with no distinct clustering by sex observed in other cell clusters. This could be explained by incomplete gene dosage compensation in the miracidium. Following investigations of differences in gene dosage compensation across developmental stages using bulk RNA-seq (Picard *et al*., 2019), our study suggests tissue-specific variation could add a further dimension to the role of dosage compensation in schistosomes. A similar phenomenon has been described in mouse embryonic stem cells and in *Drosophila*, where a precise regulation of a gene-by-gene dosage compensation mechanism is critical to the proper development of specific tissues and organs, such as the wing and eye (Valsecchi *et al.*, 2018). In contrast to the miracidium, sex differences in undifferentiated 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 (Buddenborg *et al*., 2021). In addition to using an improved reference annotation, 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 developing 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 the stem cells that persist and differentiate into the cell types of the subsequent life stages (Wang *et al*., 2018). In this respect, all miracidium somatic cells are miracidium-only cells. Our atlas shows that their transcriptional signatures are similar to comparable tissue types at later stages (Diaz *et al*., 2020; Wendt *et al*., 2020, Diaz & Attenborough *et al*., 2023) and that there may be conserved tissue-specific molecular developmental programs that are deployed multiple times during the complex life cycle. Through this atlas, we have shown 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 allow us to explore the molecular processes and pathways to reveal cell functions and interactions. In addition, high coverage data from individual cells has allowed us to examine how gene expression changes across and within tissue types. Already, a picture is emerging of how the peripheral neurons may detect environmental 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; what the contents of the penetration glands may be; and how the tegumental and stem cells are primed for the transition to the intra-molluscan 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 experimental mouse 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 fresh water (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 to fix the miracidia 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 of 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 ml of ice-cold PBS +/+ was added to bring the BSA to ~0.5%. The samples were centrifuged at 350 g for 5 minutes at 4□, and all but 20 µl of the supernatant was 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□, 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.

10X Genomics library preparation and sequencing

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 (Beckman Coulter Cytoflex S). 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) 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, 90 bp read 2, 10 bp index 1, 10 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*., 2016; 2017)) using 10X Genomics Cell Ranger (version 6.0.1) (Zheng *et al*., 2017). Before mapping, exon eight 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 gametologs. 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) (Hao *et al*., 2021) 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 gene distribution 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 practices from the authors of the software, and other similar studies; here, we were aiming to remove low-quality cells while being relatively permissive

as there may be cell types in these data outside of standard filtering parameters. Based on existing knowledge of this larva, we were expecting both multinucleate and anucleate cell type(s) as well as single-nuclei cells (Pan, 1980). 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, which 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 (producing one sorted object and one unsorted object). We used SCTransform to normalise and scale these two objects (with regression for mitochondrial percentage and sample ID). Seurat's IntegrateData function was then used to integrate the sorted and unsorted samples (with normalization.method = "SCT"). This resulted in 20,478 cells passing all the QC filters and going onto the primary analysis, 3,849 and 3,962 from the sorted samples and 6,702 and 5,965 from the unsorted samples.

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, 55 PCs 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 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 determine the identity of the cell clusters. To link gene description and chromosome locations to gene IDs, the GFF for the latest publicly available genome version was downloaded from WBPS (WBPS18 release, Howe *et al*., 2016; 2017)).

Sex-effect analysis

To investigate sex effects on the single cell 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 mapping to the WSR were classified as W+, and the remaining cells were classified as W-. The 0.2% threshold was selected (Supplementary Figure 12) to avoid classifying ZZ cells with a few transcripts erroneously mapping to a gene on the WSR 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 (using default SAM parameters), and the SAM algorithm (Tarashansky *et al*., 2019) run. For further analysis, the object was converted into a Scanpy (Wolf *et al*., (2018)) object and metadata from the Seurat analysis was attached. This method was repeated on the neuron 1 cluster alone, to better understand the transcriptome heterogeneity.

RNA velocity

All four miracidia samples and the raw data from the mother sporocyst (Diaz & Attenborough *et al*., 2023) were pseudo-mapped using kallisto bustools (v 0.27.3) (Melsted *et al*., 2021) using the lamanno workflow to the version 10 *S. mansoni* genome (WBPS18 release (Howe *et al*., 2016; 2017, with the same amendments as listed above). The filtered spliced and unspliced matrices for each sample individually were imported to R. DropletUtils (v 1.20.0) (Lun *et al*., 2019) and BUSpaRse (v 1.14.1) (Moses and Pachter, 2023) were used to process and QC the matrices, and filter out cells and genes with few reads. The matrices were then converted to Seurat objects and further filtered so only cells with the cell barcodes that passed QC filtering in the primary Seurat analysis were retained. The metadata and raw counts of the unspliced and spliced reads following filtering were exported as CSV files. The same method was performed on the mother sporocyst samples, which were combined before

exporting. For the joint analyses of miracidia and sporocyst cells, the combined sporocyst Seurat object was merged with each of the four miracidia samples separately (due to the mismatch in sample size and batch effect between sorted and unsorted samples) and all were exported.

The CSV files containing spliced and unspliced reads, together with sample metadata were combined to make an AnnData object using Scanpy (v 1.8.2, Python v 3.8.5) (Wolf *et al*., 2018). Using scvelo (v 0.2.4) (Bergen *et al*., 2020), the standard parameters were used to filter and normalise the data, compute moments, and recover dynamics. Velocity was calculated using the dynamical mode. Following PCA, a UMAP was generated (n_neighbors=10, n_pcs=40) and Leiden clustering was added (resolution=0.8). The top-ranked dynamical genes for each cluster and known genes of interest were plotted. This was performed for each of the four miracidia samples combined with the sporocyst samples, with and without the use of combat for batch correction. As the sporocyst samples were also sorted, we used the sorted miracidia samples to reduce technical effects.

To focus on the stem/germinal and tegument clusters, the same method as above was used after subsetting the data in Seurat to contain only cells from the clusters Stem/germinal, Tegument 1, and Tegument 2 in the sporocyst samples, and only Stem A-G and Tegument clusters from the miracidia clusters.

We present the stem and tegument cells from the sporocyst stage merged with sample 1 (sorted). The equivalent results from sporocysts merged with sample 4 of the miracidia and results with all cells from those two conditions are available in the supplementary data (Supplementary Figures 19-21).

EdU-staining experiment

Miracidia were collected as described above for 2 hours. As miracidia continually hatched over this time period, they ranged from 0 to 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 out, with 20 miracidia and 20 mother sporocysts per treatment examined for each replicate.

Plate-based scRNAseq analysis of individually isolated cells

A sample of thousands of miracidia was dissociated into a single cell suspension using the dissociation protocol described above. Individual gland cells, ciliary plates and unknown cells were picked up one at time using an EZ Grip transfer pipette set to aspirate a 3µl volume (Research Instruments Ltd 7-72-2800). Each cell was placed in a well of a 96-well plate containing a lysis buffer (NEBNext Cell Lysis Buffer, E6428). cDNA was generated using the NEB low input RNA kit (E6420). DNA library construction was carried out using the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (E7805). The libraries were sequenced on one lane of NovaSeq SP. The reads from the 96 samples were pseudoaligned to the version 10 *S. mansoni* genome (WBPS18 release (Howe *et al*., 2016; 2017, with the same amendments as listed above) using kallisto (v 0.44.0) (Bray *et al*., 2016)). These were imported into R (v4.3.1) and the metadata were inspected, after which only samples coming from a single cell with a minimum of 20% reads pseudoaligned were retained (n=76). The samples were then processed using sleuth (v0.30.1) (Pimentel *et al*., 2017) following guidance from sleuth tutorials, including the steps for normalising, fitting error models, and differential analysis. Differential analysis was used to identify significant genes between the cell types, which were visualised, along with known genes of interest, using sleuth's heatmap function. Furthermore, the kallisto outputs were used to calculate the top 200 genes by median transcripts per million (TPM) of the two key cell types of interest, ciliary plates and gland cells, and these were used for GO term analysis following the method described below.

***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, 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 and 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 18, WS271, Howe *et al*., 2017)(Supplementary Table 15). 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).

Opsin phylogeny

Amino acid sequences for selected metazoan opsin proteins were obtained from the GenBank database and aligned using the —auto mode of MAFFT v7.450 (Katoh & Standley, 2013), then trimmed using trimAl v1.4.rev22 (Capella-Gutierrez *et al*., 2009) to remove alignment columns with more than 50% gap characters ensuring at least 50% of alignment columns were maintained in the final alignment. Phylogenetic analysis of this alignment was performed using RaxML v8.2.12 (Stamatakis, 2014). The best-fitting empirical amino acid substitution model was identified as maximising the log-likelihood of the data under a model incorporating gamma-distributed variation in substitution rates across sites, and this optimal model (LG) was used for subsequent inference. The maximum-likelihood tree and branch lengths were identified as the highest likelihood from 25 independent searches starting from different random starting tree topologies, and support for partitions on this tree was estimated using 1000 bootstrap replicates of resampling alignment positions with replacement, using the fast bootstrap (-x) algorithm implemented in RaxML.

Protein-Protein interaction analysis

Protein 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 3-5, ciliary plates, 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 cell cluster 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

Each cluster's list of marker genes was filtered to focus on specific cluster markers. Genes were retained for analysis if they 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 as the basis for GO enrichment analysis in the Delta/Phi and Kappa stem clusters. TopGO (version 2.46.0) (Alexa & Rahnenfuhrer, 2023) 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 the statistical significance of overrepresented terms, with a FDR threshold set < 0.05.

Data (and Software) Availability

All raw sequence data is available at the ENA under study accession PRJEB45615. The scripts used to perform the analyses presented here are available at https://github.com/tessatten/singlecell-miracidia. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Funding

The project was funded by the Wellcome Trust (grant 098051 and 206194). KR also received transition funds from the Marine Biological Laboratory in Woods Hole. SRD and GR are supported by UKRI Future Leaders Fellowships [MR/T020733/1] and [MR/W013568/1], respectively.

Author Contribution

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

CD developed and optimised the dissociation protocol

KR and GR carried out the dissociation and cell collection

JG carried out cytometry

GS optimised the 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

JC carried out the opsin phylogeny

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 at the Wellcome Sanger Institute, and Karl Hoffmann, Josephine Fforde-Thomas and Benjamin Hulme for additional ongoing 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.

Figure legends

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 (larval anterior pole at the top in all images). **B)** Experimental scheme describing the parasite's life cycle (images of developmental stages not to scale), parasite dissociation, single-cell analysis of miracidium and validation pipeline. An average of 9975

miracidia per sample were dissociated; two samples were enriched for live cells (propidium iodide negative) using fluorescence-activated cell sorting (FACS), another two samples were unsorted. Cells were loaded 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 by cluster identity. **D)** Gene-expression profiles of the top two population markers identified for each cell cluster. The colours represent the expression level 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.

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 paramyosin *PRM*$^+$ and counterstaining with phalloidin reveals the nuclei of the circular body wall muscles (BWM), which 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 orthogonally to the circular muscles. Scale = 20µm. **C)** i) Segmentation of the *PRM*$^+$ cells shows 74 muscle cells in total: ~28 circular (segmented in orange), 33 longitudinal (yellow) and ii) 13 in an anterior unilateral cluster (inside dashed cyan line) adjacent to the apical gland cell (identifiable by its four nuclei, inside dashed red circle). **D)** Expression of markers for muscle 1 cluster; i) *Kunitz-type protease inhibitor* is expressed in circular BWM and *wnt-11-1* is expressed in 7 *Kunitz-type protease inhibitor*$^+$ cells at the posterior pole; ii) close-up of posterior end of another miracidium showing the co-expression of *Kunitz-type protease inhibitor* 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 13 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).

Figure 3. Neural complexity and ciliary plates in a simple larva.

A) Expression profiles of cell marker genes that are specific or enriched in the five neuronal clusters and ciliary plates. Genes validated by *in situ* hybridisation (ISH) are marked in red. *Complexin* is expressed in all five nerve 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 209 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 of *cpx*$^+$ cells 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 2 (Smp_201600) and neuron 3 (Smp_071050) shows expression in multiple cells of the brain, including cells adjacent to each other. Smp_201600 is also expressed in the apical gland cell (red dashed circle). **D)** The top marker for neuron 4 Smp_319480 is expressed in four cells anteriorly (arrows) 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 **(iii)** seen here expressing the ciliary plate marker gene Smp_096390. **E)** Neuron 5 marker Smp_343710, an EF-hand domain-containing protein, is expressed in **(i)** 10-20 cells whose nuclei sit outside of, and anterior to, the brain, in and around the paired lateral glands and their secretory ducts, and **(Eii & iii, F)** in a pair of bilaterally symmetrical bulbous protrusions. **F)** Summary schematic of the *in situ* expression of marker genes for the neural cell clusters. **G)(i)** A top marker for ciliary plates, a calcium binding protein (Smp_096390), shows transcripts expressed in all the ciliary plates of all four tiers (brackets) 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).

Figure 4. Identification of the tegument, protonephridia, parenchymal cells in the miracidium. **A)** Expression profiles of cell marker genes specific or enriched in these cell type clusters. Genes validated by *in situ* hybridisation are in red. **Bi & ii)** WISH and segmentation of marker genes show that 6 cells express Smp_335600 (*ShKT-domain protein*, a marker for protonephridia), 46 cells that express *Meg-6* (a marker for tegument), and seven cells that express an uncharacterised gene, Smp_318890 (a marker for parenchymal cells). **C**

& D) 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 **Dii & iii** expressing the ciliary plate marker Smp_096390) (asterisks). **E)** The protonephridial marker Smp_335600 shows 'S'-shaped expression with transcripts extending from the nucleus of the anterior flame cell (AFC) along the excretory tubule and its nucleus (ETN) to the posterior flame cell (PFC), **Eii)** it is expressed around the nuclei (arrow) rather than the barrels (arrowhead) of the flame cells. **Fi)** The pan-parenchymal marker, Smp_318890, was expressed in 2 anterior cells (one on either side of the brain) and 5-7 cells posterior to the brain. **Fii)** These cells have long cytoplasmic protrusions that reach between all the other cells, **Fiii)** including the *ago2-1*$^{+}$ stem cells.

Figure 5. Two defined populations of stem cells cluster by sex. A) Expression profiles of cell marker genes that are specific or enriched in the stem cell clusters. Genes specific to the W (female-specific) sex chromosome 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 reveal there are more *p53*$^{+}$ cells than *UPPA*$^{+}$ cells (in this larva,15 are p53^{+} and 9 are UPPA^{+}).

Figure 6. RNA velocity analysis of stem and tegument cells from the miracidium and sporocyst show lineage-specific gene dynamics. A) UMAP shows the life cycle stage origin of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field show the generalized direction of RNA velocity. **C)** Latent time analysis shows an estimated temporal-relationship between cells. **B)** and **C)** are based on the expression of both spliced and unspliced transcripts, and their expression dynamics across cells and genes. The phase plot, velocity, and expression were calculated for **D)** *p53-1*, **E)** *Zfp-1*, **F)** *p-53-2* and **G)** *eled*

(no batch correction). For all of **D) - G)**, the phase plot (left plot) shows the proportion of spliced and unspliced transcripts in each cell, where each point is a cell and is coloured by the clusters in **B)**. The purple almond-shape overlaid represents the processes of transcription, splicing, and degradation, where this can be modelled. The dashed line shows the estimated steady state where RNA transcription is constant. The middle panel shows the RNA velocity, which for each gene is based on how the observations deviated from the estimated steady state towards induction or repression. The right panel shows gene expression. *p53-1* and *zfp-1* are predominantly expressed in the Delta/Phi-like miracidia stem cells and some miracidia tegument, and velocity indicates active expression of *p53-1* but downregulation of *zfp-1* in the stem cells. *p53-2* is most highly expressed in the Kappa-like miracidia stem cells, and velocity indicates this gene is being actively transcribed. *eled* expression is very low and only spliced transcripts have been detected, but is generally restricted to the Kappa-like miracidia stem cells.

Figure 7. **A model for the fate of the two stem populations in the miracidium**. Adding single-cell data for miracidia stem cells (this study) to existing stem cell scenarios on *Schistosoma mansoni* development (Wang *et al*., 2018; Li *et al*., 2021) shows the continuum of the Kappa (κ) population from the miracidium, through the intra-molluscan stages to the juvenile and adult stages inside the mammalian host. Wang *et al*. (2018) proposed that the κ cells give rise to epsilon (□), *eled*$^{+}$, cells in the juvenile primordial testes, ovaries, and vitellaria (germline), as well as in a gradient increasing toward the posterior growth zone (soma). They suggested that germ cells may be derived from ε-cells early in juvenile development, and *eled* is the earliest germline marker yet identified in schistosomes. This led to the idea that *S. mansoni* does not specify its germline until the juvenile stage (Wang *et al.*, 2018) and a germline-specific regulatory program (including *eled, oc-1, akkn, nanos-1, boule*) was identified in intra-mammalian stages (Wang *et al*., 2018; Li *et al.*, 2021). We show expression of these genes in κ stem cells in the miracidium. This suggests that after ~6 days of embryogenesis, at hatching of the miracidium, the cells that may contribute to the germline might already be segregated into the κ population and the molecular regulatory program that differentiates somatic (delta/phi, □/□) and germ cell (κ) lineages is active. Furthermore, as *p53-2* plays a genotoxic stress response role in adult reproductive cells (Wendt *et al*., 2022), it's expression in κ cells in the miracidia is another line of evidence that indicates that this population may contain the pluripotent stem cells that likely give rise to the germline.

Figure 8. **Tissue level segmentation of a miracidium reveals the contribution of each tissue to this simple larva**. *In situ* hybridization using HCR (hybridization chain reaction) enabled multiple tissue-level marker genes to be visualised simultaneously in the same larva. The nuclei of the cells of each tissue type were manually segmented using TrakEM2 in ImageJ (Cardona *et al*., 2012).

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 five 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 cell cluster 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)** 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. **D)** Neuron 4 marker Smp_319480 (putative mTORC1) transcripts are expressed at the base of these cilia; **E)** and some transcripts also reach the nuclei of circular body wall muscles. **F)** The expression of *p25 alpha* (which is a marker for ciliated cells in adult worms, Wendt *et al*., 2020) in the submuscular cells suggests that they may be ciliated.

Supplementary Figure 3. **Phylogeny of metazoan opsins.** There are four *Schistosoma mansoni* opsins that are expressed in two sub-clusters of neuron 1; our analysis shows that three belong to the rhabdomeric opsin clade and the fourth is a peropsin. Scale bar shows branch lengths in expected amino acid substitutions per site. Shading of circles at internal nodes shows percentage support for the partition implied by that node from 1,000 non-parametric bootstrap samples.

Supplementary Figure 4. **The expression of two Polycystin genes and a Sox transcription factor in neuron 4 cells. A)** Polycystin 1 and 2 are co-expressed in neuron 4 cells, orthologs in other animals form heterodimers in ciliated cells that function to detect flow sheer and vibrations (Kim *et al*., 2016; Calderon *et al*., 2018). **B)** In planaria, a transcription factor Sox B1-2 regulates polycystin genes and together function in sensory

neurons that detect flow. The co-expression of polycystin genes with a Sox transcription factor in neuron 4 cells indicates that this transcription factor may regulate these polycystin genes in the miracidium.

Supplementary Figure 5. STRING analysis of neural clusters and ciliary plates. STRING analysis of the top 100 marker genes for each of the neural populations whose predicted networks have functional enrichment. **A)** Neuron 4, **B)** Neuron 5, **C)** Ciliary plate, **D)** Neuron 3. Neuron 1 is not a defined cluster and Neuron 2 showed no functional enrichment. Lines (edges) connecting nodes are based on evidence of the function of homologs. 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. Gene expression by transcripts per million in the ciliary plates and gland cells. Heatmap showing the gene expression, by transcripts per million, detected in the ciliary plates and gland cells hand-picked for plate-based scRNA-seq. Genes included; the top 20 significant genes between gland cell and ciliary plates, two neural markers, five VAL genes, five significant genes between gland cells and other cells, eight genes previously reported to produce secreted proteins in the miracidia (Wang *et al*., 2016; Wu *et al*., 2006), five top marker genes from the ciliary plate cluster.

Supplementary Figure 7. STRING analysis of tegument and protonephridia clusters. STRING analysis of the top 100 marker genes for **A)** tegument, **B)** protonephridia. Lines (edges) connecting nodes are based on evidence of the function of homologs. 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 8. 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 homologs. 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 9. Gene ontology (GO) analysis of marker genes identified in the miracidia cell clusters. GO analysis was performed on each cluster using genes with a minimum AUC score of 0.7. Enriched biological process (BP) GO terms are shown for each cluster, and have been filtered to only show terms supported by a minimum of two genes. Top panels (left to right) show parenchyma, muscle, and tegument clusters, middle panels (left to right) show ciliary plate, neuron, and protonephridia clusters, and the bottom panel shows stem clusters. Full list of GO terms available in Supplementary Table 6.

Supplementary Figure 10. 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 11 - The effect of sex-linked genes on cell clustering. *Schistosoma mansoni* has 7 pairs of autosomes and one pair of sex chromosomes; males have ZZ and females have ZW (Buddenborg *et al*., 2021). The sex chromosomes are composed of sex-specific regions that are unique to each chromosome flanked by pseudoautosomal regions 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 gametologs (genes with homologous copies on the WSR and ZSR). The Z-specific region (ZSR) contains 941 genes, of which only 35 are gametologs. **A & B)** The proportion of reads mapping to the WSR and ZSR across all cell type clusters shows **A)** putative male cells (Stem A and C 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 any of the 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. In all four scenarios the stem cell clusters

remained separated by sex. ZSR or WSR expression is therefore sufficient on their own to split these stem clusters in two.

Supplementary Figure 12. Classification of cells based on percentage of reads mapping to W-specific region indicates incomplete dosage compensation of Z-specific region 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.

Supplementary Figure 13. EdU labelling shows that there are no dividing cells during the free-swimming miracidia stage. **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, three day EdU pulse on freshly-transformed mother sporocysts shows many dividing cells incorporating EdU, all treatments n = 30.

Supplementary Figure 14. 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 15. Differential gene expression between the two stem cell populations, Delta-phi and Kappa. **A)** Heatmap of top differentially expressed genes (by adjusted p-value) from each of Delta/Phi and Kappa populations show a 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 the structural constituent of the ribosome and translation,

while up-regulated genes in Kappa were related to transcription and metabolism**.**

Supplementary Figure 16. RNA velocity analysis of stem and tegument cells from the miracidium and sporocyst (with batch correction) show lineage-specific gene dynamics. A) UMAP shows the life cycle stage origin of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field shows generalised direction of RNA velocity. **C)** Latent time analysis shows an estimated temporal-relationship between cells. The phase plot, velocity, and expression were calculated for **D)** *p53-1*, **E)** *Zfp-1*, **F)** *p-53-2* and **G)** *eled*. *p53-1* and *zfp-1* are predominantly expressed in the Delta/Phi-like miracidia stem cells and miracidia tegument, and velocity indicates active expression of *p53-1* but downregulation of *zfp-1* in the stem cells. *p53-2* is most highly expressed in the Kappa-like miracidia stem cells, and velocity indicates this gene is being actively transcribed. *eled* expression is very low and only spliced transcripts have been detected, but is generally restricted to the Kappa-like miracidia stem cells.

Supplementary Figure 17. RNA velocity analysis of stem and tegument cells from the miracidium and sporocyst (without batch correction). A) UMAP shows the life cycle stage origin of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field shows generalised direction of RNA velocity. Phase plot, velocity and expression as calculated for genes mentioned in text: blue outline = known molecular regulators of tegument development and tegument markers in adult *S. mansoni* and planaria. Green outline = miracidia and sporocyst tegument marker genes. Red outline = genes revealed by RNA velocity analysis to be highly dynamic in the Delta/Phi, miracidia and sporocyst tegument cells.

Supplementary Figure 18. RNAvelocity analysis of stem and tegument cells from the miracidium and sporocyst (without batch correction). A) UMAP shows the life cycle stage origin of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field shows generalised direction of RNA velocity. Phase plot, velocity and expression as calculated for genes that are mentioned in the text: red outline = known molecular regulators of germline stem cells in adult *S. mansoni*. Blue outline = pluripotency gene. *Oc-1* expression is low with little velocity signal, but the expression is predominantly seen in the Kappa-like miracidia stem cells (and Stem G). *Irx* is also primarily expressed in the Kappa-like miracidia stem cells with velocity results indicating an accumulation of spliced

transcripts. *Irx* transcripts are also detected in Stem G and the sporocyst stem/germinal cells. *Akkn* expression is low and largely seen in the Kappa-like and Stem G miracidia stem clusters. Lsm14-domain containing protein transcripts are most consistently expressed in the Kappa-like miracidia stem cells. *Klf* is most highly expressed in sporocyst stem/germinal cells, and the Kappa-like miracidia stem cluster.

Supplementary Figure 19. RNA velocity analysis of all tissue types from the miracidium and sporocyst show lineage-specific gene dynamics. A) UMAP shows the life cycle stage origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis shows an estimated temporal-relationship between cells. Phase plot, velocity and expression as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst cells combined with miracidia sample 1 cells). *P53-1* is most highly expressed in the Delta/phi-like miracidia stem clusters and some of the miracidia tegument cells, and the velocity indicates active transcription. *Zfp-1* is expressed most highly in the same clusters, but velocity suggests that the majority of transcripts are spliced. *P53-2* expression is highest in the Kappa-like miracidia stem clusters, but the expression is detected in several other clusters, including muscle, neuron, and other stem clusters. Velocity indicates that this gene is being actively transcribed in the majority of these clusters. *Eled* expression is very low, but restricted to the Kappa-like miracidia stem clusters.

Supplementary Figure 20. RNA velocity analysis of all tissue types from the miracidium and sporocyst shows the dynamics of described germline-related genes. A) UMAP shows the life cycle stage origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity analysis flow field shows the generalized direction of RNA velocity. Phase plot, velocity and expression as calculated for **E)** *oc-1*, **F)** *boule*, **G)** *irx*, **H)** *akkn*, and **I)** an lsm14-domain-containing protein. (no batch correction, sporocyst cells combined with miracidia sample 1 cells.). *Oc-1* expression is low but generally specific to the two Kappa-like miracidia stem cells, and Stem G. Only spliced *Boule* transcripts were detected, and with low expression, predominantly in the Stem B and Stem E clusters. *Irx* was most highly expressed in Kappa-like miracidia stem clusters, as well as Stem E, Stem G, and sporocyst stem/germinal cells. *Akkn* is also largely expressed in the Kappa-like miracidia stem clusters. The lsm14-domain-containing protein expression is low, but seen most in Kappa-like miracidia stem clusters, Stem G, and several neural clusters.

Supplementary Figure 21. RNAvelocity analysis of all tissue types from the miracidium and sporocyst show lineage-specific gene dynamics. A) UMAP shows the life cycle stage origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis shows an estimated temporal-relationship between cells. Phase plot, velocity and expression as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst cells combined with miracidia sample 4 cells). *p53-1* and *zfp-1* are predominantly in the Delta/Phi-like miracidia stem cell clusters (Stem C and D) and also in the miracidia tegument. Velocity indicates that *p53-1* is being induced in the stem cells and repressed in the tegument, while *zfp-1* shows a gradient of velocity in the stem cells and repression in the tegument. *P53-2* is most highly expressed in the Kappa-like miracidia stem cells (Stem A and B) and Stem F, with velocity results indicating active transcription. *eled* expression was very low, but primarily in the Kappa-like miracidia stem cells (particularly Stem A).

References

1. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Development. 2017;144: 3625–3632. doi:10.1242/dev.151142

2. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. In: R package version 2.54.0 [Internet]. 2023. doi:10.18129/B9.bioc.topGO

3. Ashton PD, Harrop R, Shah B, Wilson RA. The schistosome egg: development and secretions. Parasitology. 2001;122: 329–338. doi:10.1017/s0031182001007351

4. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38: 1408–1414. doi:10.1038/s41587-020-0591-3

5. Bezares-Calderón LA, Berger J, Jasek S, Verasztó C, Mendes S, Gühmann M, et al. Neural circuitry of a polycystin-mediated hydrodynamic startle response for predator avoidance. Elife. 2018;7. doi:10.7554/eLife.36262

6. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

7. Buddenborg SK, Lu Z, Sankaranarayan G, Doyle SR, Berriman M. The stage- and sex-specific transcriptome of the human parasite Schistosoma mansoni. Sci Data. 2023;10: 775. doi:10.1038/s41597-023-02674-2

8. Buddenborg SK, Tracey A, Berger DJ, Lu Z, Doyle SR, Fu B, et al. Assembled chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the evolution of its ZW sex-determination system. bioRxiv. 2021; 2021.08.13.456314-2021.08.13.456314. doi:10.1101/2021.08.13.456314

9. Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, et al. TrakEM2 software for neural circuit reconstruction. PLoS One. 2012;7: e38011–e38011. doi:10.1371/journal.pone.0038011

10. Chernin E. Interference with the capacity of Schistosoma mansoni miracidia to infect the molluscan host. J Parasitol. 1968;54: 509–516.

11. Choi HMT, Calvert CR, Husain N, Huss D, Barsi JC, Deverman BE, et al. Mapping a multiplexed zoo of mRNA expression. Development. 2016;143: 3632–3637. doi:10.1242/dev.140137

12. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development. 2018;145. doi:10.1242/dev.165753

13. Collins JJ 3rd, 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 Negl Trop Dis. 2011;5: e1009–e1009. doi:10.1371/journal.pntd.0001009

14. Costain AH, MacDonald AS, Smits HH. Schistosome Egg Migration: Mechanisms, Pathogenesis and Host Immune Responses. Front Immunol. 2018;9: 3042. doi:10.3389/fimmu.2018.03042

15. Diaz Soria CL, Attenborough T, Lu Z, Graham J, Hall C, Thompson S, et al. Single cell transcriptomics of the human parasite Schistosoma mansoni first intra-molluscan stage reveals tentative tegumental and stem cell regulators. bioRxiv. 2023; 2023.05.02.538881-2023.05.02.538881. doi:10.1101/2023.05.02.538881

16. Diaz Soria CL, Lee J, Chong T, Coghlan A, Tracey A, Young MD, et al. Single-cell atlas of the first intra-mammalian developmental stage of the human parasite Schistosoma mansoni. Nat Commun. 2020;11: 6411. doi:10.1038/s41467-020-20092-5

17. Gąsiorowski L, Andrikou C, Janssen R, Bump P, Budd GE, Lowe CJ, et al. Molecular evidence for a single origin of ultrafiltration-based excretory organs. Curr Biol. 2021;31: 3629-3638.e2. doi:10.1016/j.cub.2021.05.057

18. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184: 3573-3587.e29. doi:10.1016/j.cell.2021.04.048

19. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016;44: D774-80. doi:10.1093/nar/gkv1217

20. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215: 2–10. doi:10.1016/j.molbiopara.2016.11.005

21. IHME. 2019 Global Burden of Disease (GBD) study [Internet]. 2019. Available: https://vizhub.healthdata.org/gbd-results?params=gbd-api-2019-permalink/d30124bc2eaad3d977583362d4d37591 [vizhub.healthdata.org]

22. Jurberg AD, Gonçalves T, Costa TA, de Mattos ACA, Pascarelli BM, de Manso PPA, et al. The embryonic development of Schistosoma mansoni eggs: proposal

for a new staging system. Dev Genes Evol. 2009;219: 219–234. doi:10.1007/s00427-009-0285-9

23. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang T-H, Liu Y, et al. Notum deacylates Wnt proteins to suppress signalling activity. Nature. 2015;519: 187–192. doi:10.1038/nature14259

24. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods. 2014;11: 740–742. doi:10.1038/nmeth.2967

25. Kim S, Nie H, Nesin V, Tran U, Outeda P, Bai C-X, et al. The polycystin complex mediates Wnt/Ca2+ signalling. Nat Cell Biol. 2016;18: 752–764. doi:10.1038/ncb3363

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

27. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560: 494–498. doi:10.1038/s41586-018-0414-6

28. Li P, Nanes Sarfati D, Xue Y, Yu X, Tarashansky AJ, Quake SR, et al. Single-cell analysis of Schistosoma mansoni identifies a conserved genetic program controlling germline stem cell fate. Nat Commun. 2021;12: 485. doi:10.1038/s41467-020-20794-w

29. 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;17: 329–340. doi:10.1016/j.stem.2015.07.002

30. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC, et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. 2019;20: 63. doi:10.1186/s13059-019-1662-y

31. Mair GR, Maule AG, Day TA, Halton DW. A confocal microscopical study of the musculature of adult Schistosoma mansoni. Parasitology. 2000;121 (Pt 2: 163–170. doi:10.1017/s0031182099006174

32. Maldonado JF, Acosta-Matienzo J. Biological Studies on the Miracidium of Schistosoma Mansoni. Am J Trop Med. Arlington VA, USA: The American Society of Tropical Medicine and Hygiene; 1948;s1-28: 645–657. doi:10.4269/ajtmh.1948.s1-28.645

33. Mann VH, Morales ME, Rinaldi G, Brindley PJ. Culture for genetic manipulation of developmental stages of Schistosoma mansoni. Parasitology. 2010;137: 451–462. doi:10.1017/S0031182009991211

34. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019;8: 329-337.e4. doi:https://doi.org/10.1016/j.cels.2019.03.003 [doi.org]

35. McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou X-N. Schistosomiasis. Nat Rev Dis Prim. 2018;4: 13. doi:10.1038/s41572-018-0013-8

36. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KH (Joseph), et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat Biotechnol. 2021;39: 813–818. doi:10.1038/s41587-021-00870-2

37. Moses L LP. BUSpaRse: kallisto | bustools R utilities [Internet]. 2023. doi:10.18129/B9.bioc.BUSpaRse

38. Nanes Sarfati D, Li P, Tarashansky AJ, Wang B. Single-cell deconstruction of stem-cell-driven schistosome development. Trends Parasitol. 2021;37: 790–802. doi:https://doi.org/10.1016/j.pt.2021.03.005 [doi.org]

39. Orosz F, Ovádi J. TPPP orthologs are ciliary proteins. FEBS Lett. 2008;582: 3757–3764. doi:10.1016/j.febslet.2008.10.011

40. Pan SC. The fine structure of the miracidium of Schistosoma mansoni. J Invertebr Pathol. 1980;36: 307–372. doi:https://doi.org/10.1016/0022-2011(80)90040-3 [doi.org]

41. Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body axis. Cell. 2009;139: 1056–1068. doi:10.1016/j.cell.2009.11.035

42. Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, et al. Dosage Compensation throughout the Schistosoma mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome. Genome Biol Evol. 2019;11: 1909–1922. doi:10.1093/gbe/evz133

43. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14: 687–690. doi:10.1038/nmeth.4324

44. Scimone ML, Cote LE, Reddien PW. Orthogonal muscle fibres have different instructive roles in planarian regeneration. Nature. 2017;551: 623–628. doi:10.1038/nature24660

45. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell. 2015;17: 360–372. doi:https://doi.org/10.1016/j.stem.2015.07.013 [doi.org]

46. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, et al. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science (80-). American Association for the Advancement of Science; 2019;365: eaav9314–eaav9314. doi:10.1126/science.aav9314

47. Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, Reischl R, Taudes E, Lassnig M, Technau U. Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages,Cell Reports, 40, 12,2022, 111370, https://doi.org/10.1016/j.celrep.2022.111370.

48. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47: D607–D613. doi:10.1093/nar/gky1131

49. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell. 2007;131: 861–872. doi:https://doi.org/10.1016/j.cell.2007.11.019 [doi.org]

50. Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. Shalek AK, Barkai N, editors. Elife. eLife Sciences Publications, Ltd; 2021;10: e66747–e66747. doi:10.7554/eLife.66747

51. Tarashansky AJ, Xue Y, Li P, Quake SR, Wang B. Self-assembling manifolds in single-cell RNA sequencing data. Shalek AK, Barkai N, Yanai I, editors. Elife. eLife Sciences Publications, Ltd; 2019;8: e48994–e48994. doi:10.7554/eLife.48994

52. 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;9: 3626. doi:10.1038/s41467-018-05642-2

53. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2023; doi:10.1038/s41587-023-01773-0

54. Verasztó C, Jasek S, Gühmann M, Shahidi R, Ueda N, Beard JD, et al. Whole-animal connectome and cell-type complement of the three-segmented Platynereis dumerilii larva. bioRxiv. 2020; 2020.08.21.260984-2020.08.21.260984. doi:10.1101/2020.08.21.260984

55. Vergara HM, Pape C, Meechan KI, Zinchenko V, Genoud C, Wanner AA, et al. Whole-body integration of gene expression and single-cell morphology. Cell. 2021;184: 4819-4837.e22. doi:https://doi.org/10.1016/j.cell.2021.07.017 [doi.org]

56. Vogg MC, Owlarn S, Pérez Rico YA, Xie J, Suzuki Y, Gentile L, et al. Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. Dev Biol. 2014;390: 136–148. doi:https://doi.org/10.1016/j.ydbio.2014.03.016 [doi.org]

57. Wang B, Collins III JJ, Newmark PA. Functional genomic characterization of neoblast-like stem cells in larval Schistosoma mansoni. Sánchez Alvarado A, editor. Elife. eLife Sciences Publications, Ltd; 2013;2: e00768–e00768. doi:10.7554/eLife.00768

58. Wang B, Lee J, Li P, Saberi A, Yang H, Liu C, et al. Stem cell heterogeneity drives the parasitic life cycle of Schistosoma mansoni. Sánchez Alvarado A, editor. Elife. eLife Sciences Publications, Ltd; 2018;7: e35449–e35449. doi:10.7554/eLife.35449

59. Wang T, Zhao M, Rotgans BA, Strong A, Liang D, Ni G, et al. Proteomic Analysis of the Schistosoma mansoni Miracidium. PLoS One. Public Library of Science; 2016;11: e0147247–e0147247.

60. Wendt GR, Collins JNR, Pei J, Pearson MS, Bennett HM, Loukas A, et al. Flatworm-specific transcriptional regulators promote the specification of tegumental progenitors in Schistosoma mansoni. Patel NH, editor. Elife. eLife Sciences Publications, Ltd; 2018;7: e33221–e33221. doi:10.7554/eLife.33221

61. Wendt GR, Shiroor DA, Adler CE, Collins JJ. Convergent evolution of a genotoxic stress response in a parasite-specific p53 homolog. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2022;119: e2205201119–e2205201119. doi:10.1073/pnas.2205201119

62. Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, et al. A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood feeding. Science (80-). 2020;369. doi:10.1126/science.abb7709

63. WHO. Schistosomiasis [Internet]. 2023. Available: https://www.who.int/news-room/fact-sheets/detail/schistosomiasis [who.int]

64. Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW. Muscle Cells Provide Instructions for Planarian Regeneration. Cell Rep. 2013;4: 633–641. doi:https://doi.org/10.1016/j.celrep.2013.07.022 [doi.org]

65. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19: 15. doi:10.1186/s13059-017-1382-0

66. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8: 14049. doi:10.1038/ncomms14049

Supplementary Video legends

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

Supplementary Video 2 – Segmentation of *paramyosin* (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 *Kunitz-type protease inhibitor* (Smp_052230) (cyan) and *Wnt-11-1* (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-11-1* (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 2 (Smp_201600) (magenta) and neuron 3 (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 ciliary plate 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, and 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 ciliary plate 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 4 marker Smp_319480 (magenta) and ciliary plate marker *16 kDa calcium binding protein* (Smp_096390) (green), counterstained with DAPI (white), showing that the clusters of neuron 4 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 *ShKT-domain containing protein* (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.

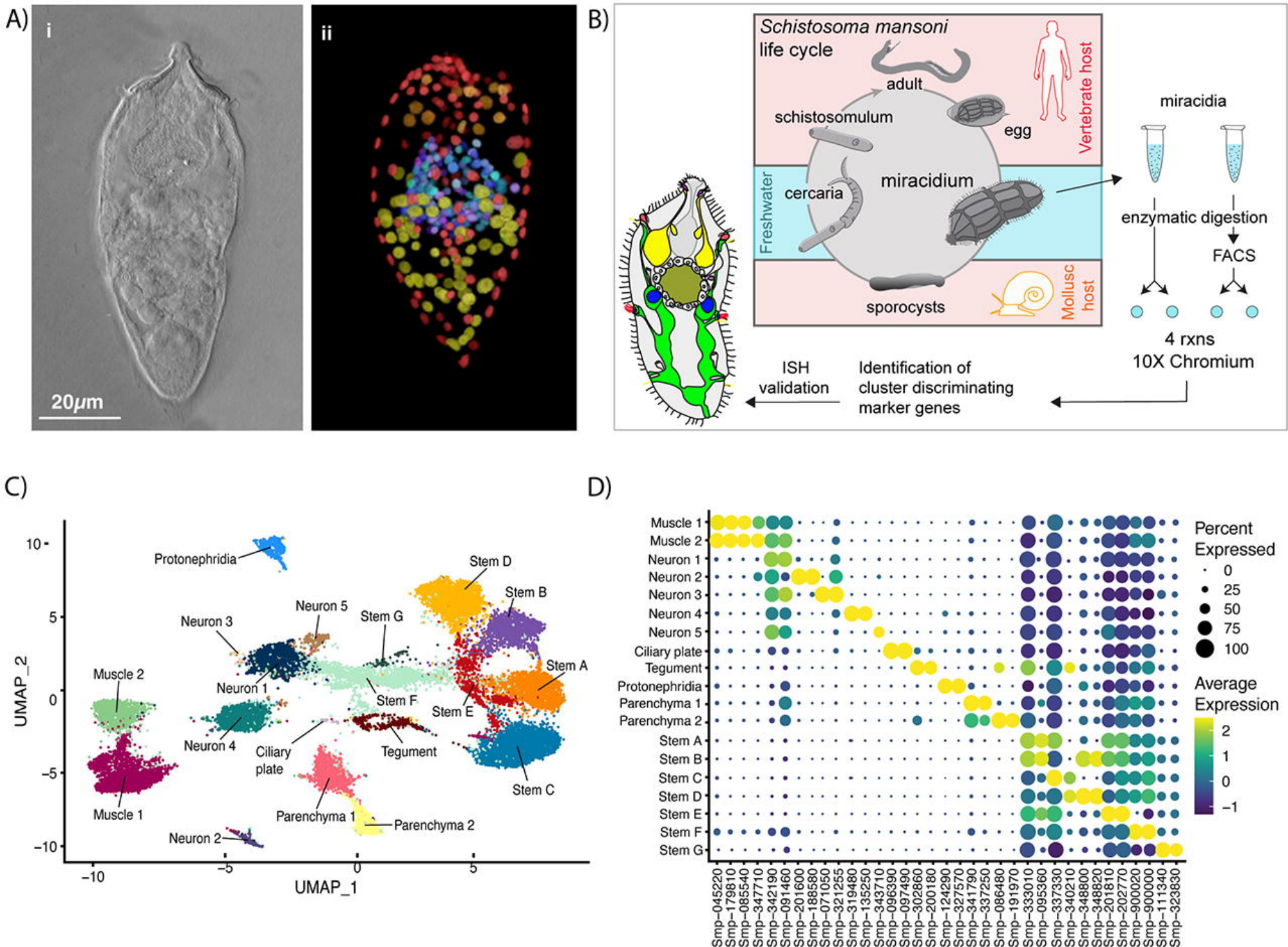
Supplementary Video 13 - Confocal z-stack of WISH of markers for parenchyma 1 and 2 (Smp_318890) (green) and stem cells *ago2-1* (Smp_179320) (magenta), counterstained with DAPI (white), shows the long cytoplasmic protrusions of the parenchymal cells reaching between the stem and other cells.

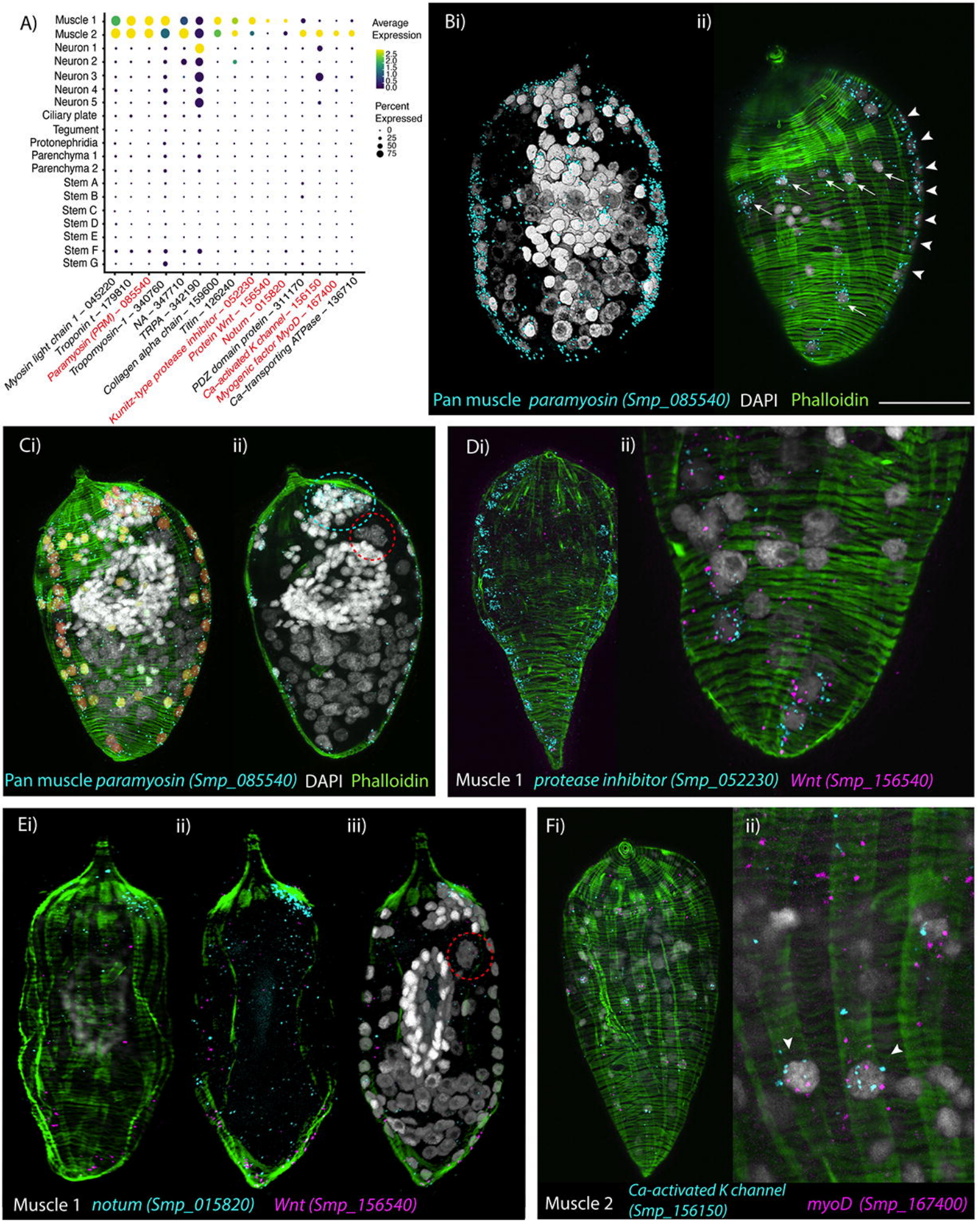
Supplementary Video 14 - Maximum intensity projection from confocal z-stack of WISH of marker for parenchyma 1 and 2 (Smp_318890) (green) counterstained with DAPI (white), shows the distribution of the parenchymal cells and their reach in the intercellular space of the larva.

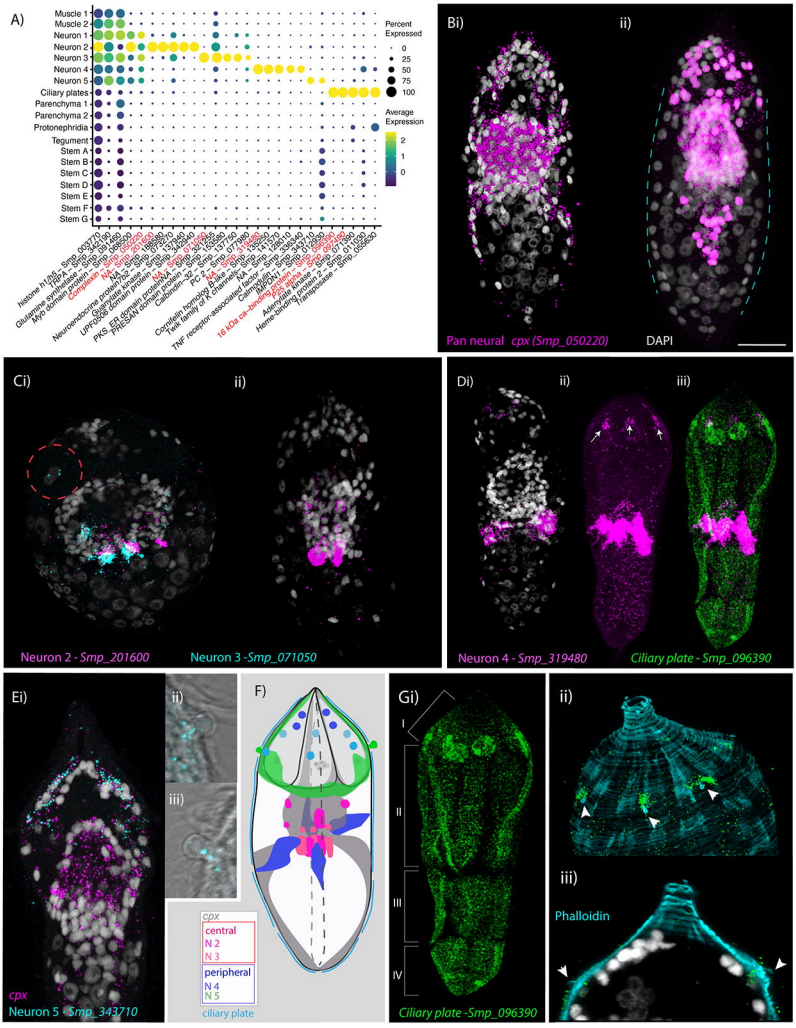
Supplementary Video 15 - Confocal z-stack of WISH of markers for tegument *Meg6* (Smp_163710) (magenta) and *dynein light chain* (Smp_200180) (cyan), counterstained with phalloidin (green) and DAPI (white), shows the nuclei sit below the body wall muscle, and cytoplasmic protrusions reach between muscle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

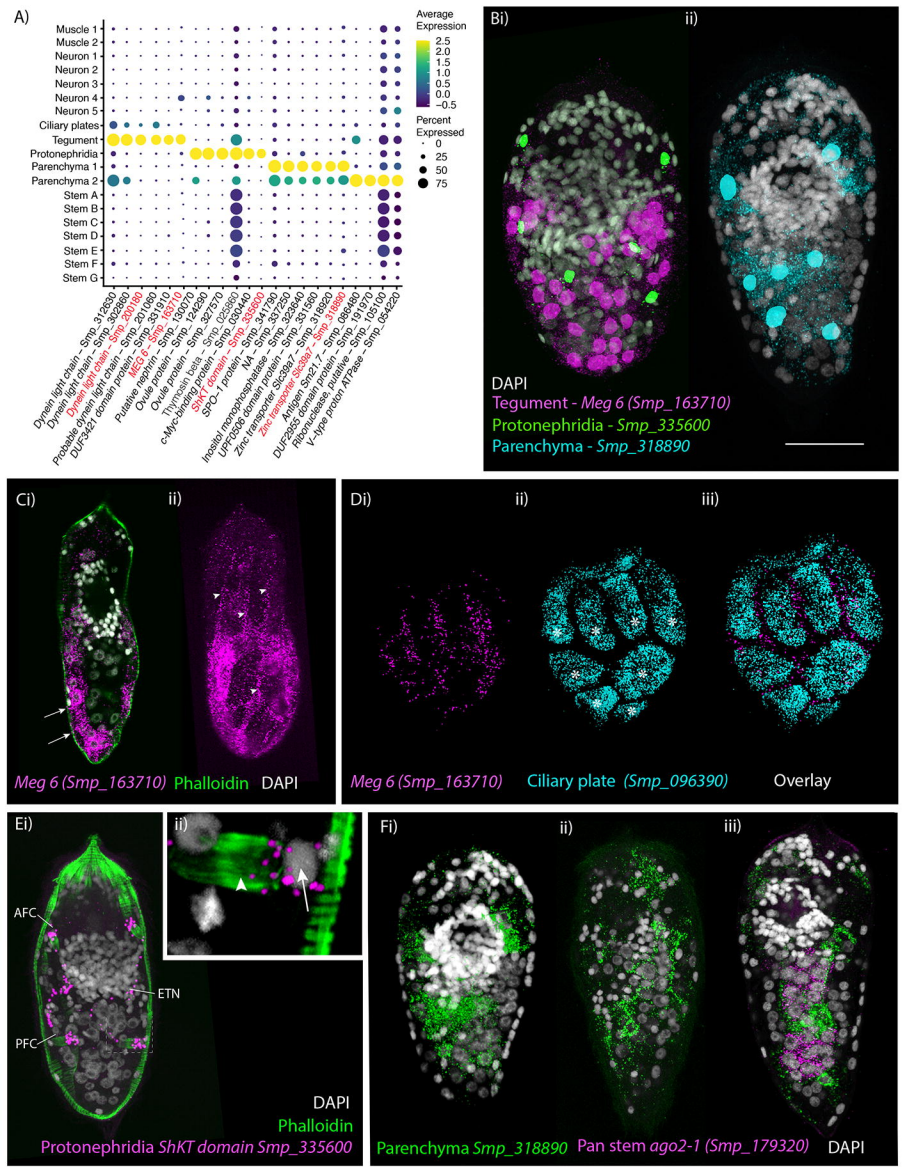
Supplementary Video 16 - Confocal z-stack of WISH of markers for stem clusters Delta/Phi *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (green) and stem E *Phosphomevalonate kinase* (Smp_072460) (magenta), counterstained with DAPI (white), shows there are no cells that are *pmk*-positive but negative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

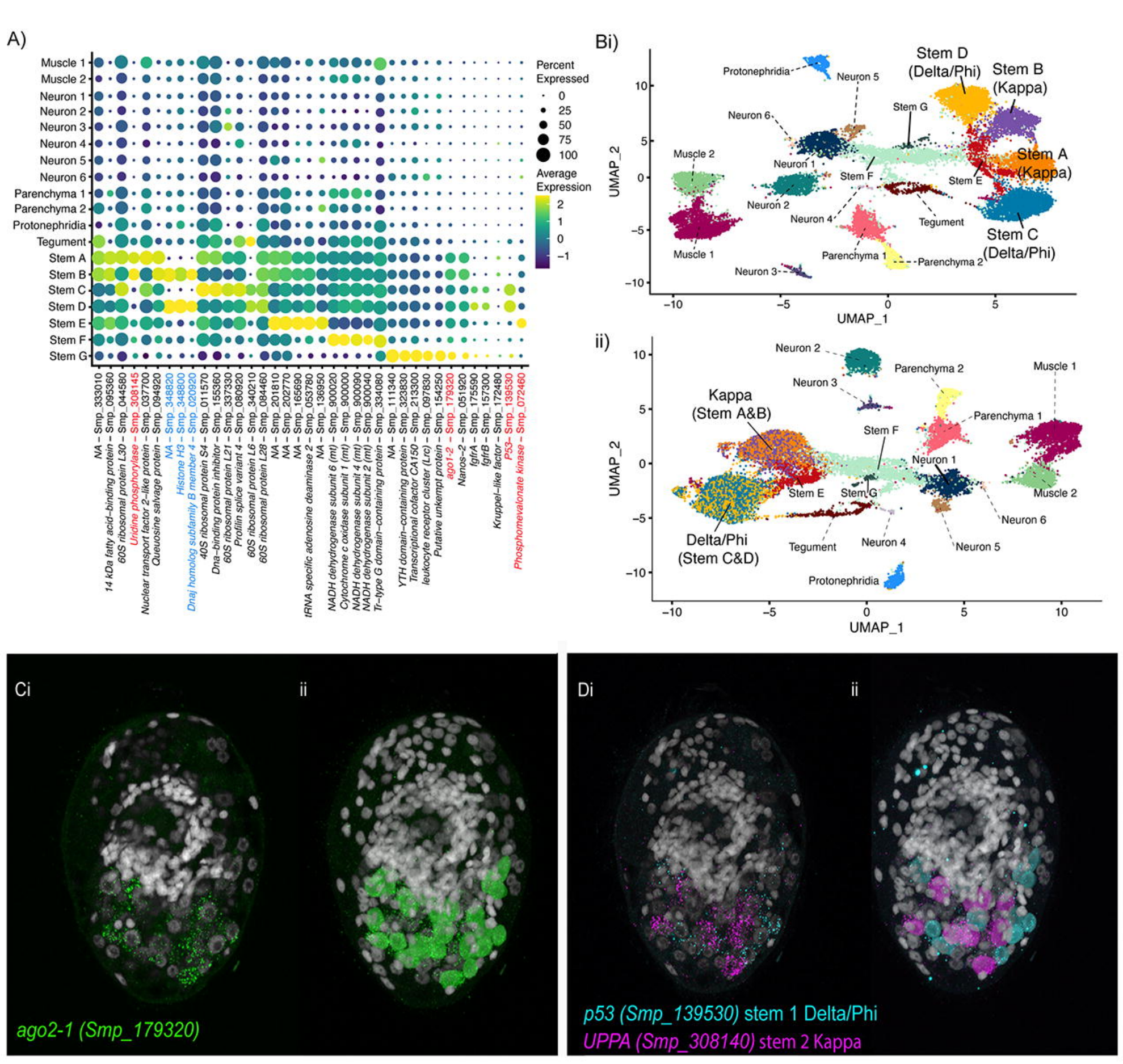
Supplementary Video 17 - Confocal z-stack of WISH of markers for stem clusters Delta/Phi *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (magenta) and the pan-stem marker *ago2-1* (Smp_179320) (green), counterstained with DAPI (white), shows there are no cells that are *ago2-1* positive but negative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

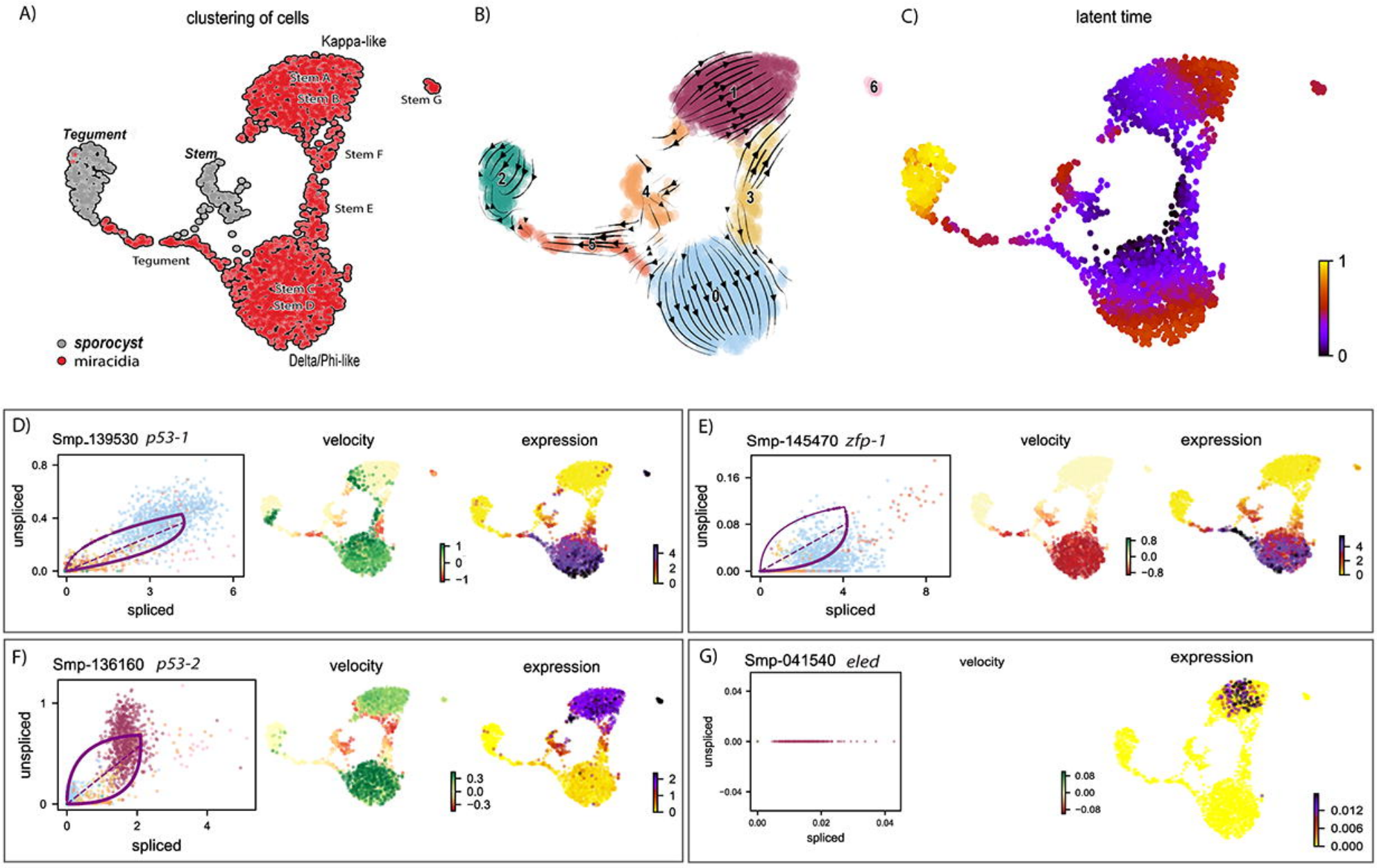

A)
i
ii
20μm
B)
Schistosoma mansoni
life cycle
adult
schistosomulum
egg
Vertebrate host
Freshwater
cercaria
miracidium
sporocysts
Mollusc host
miracidia
enzymatic digestion
FACS
4 rxns
10X Chromium
Identification of
cluster discriminating
marker genes
ISH
validation
C)
Protonephridia
Stem D
Stem B
Neuron 5
Neuron 3
Stem G
Muscle 2
Stem A
Neuron 1
Stem F
Stem E
Neuron 4
Ciliary
plate
Tegument
Stem C
Muscle 1
Parenchyma 1
Parenchyma 2
Neuron 2
UMAP_2
UMAP_1
D)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plate
Tegument
Protonephridia
Parenchyma 1
Parenchyma 2
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Percent
Expressed
0
25
50
75
100
Average
Expression
2
1
0
−1
Smp−045220
Smp−179810
Smp−085540
Smp−347710
Smp−342190
Smp−091460
Smp−201600
Smp−188580
Smp−071050
Smp−321255
Smp−319480
Smp−135250
Smp−343710
Smp−096390
Smp−097490
Smp−302860
Smp−200180
Smp−124290
Smp−327570
Smp−341790
Smp−337250
Smp−086480
Smp−191970
Smp−333010
Smp−095360
Smp−337330
Smp−340210
Smp−348800
Smp−348820
Smp−201810
Smp−202770
Smp−900020
Smp−900000
Smp−111340
Smp−323830

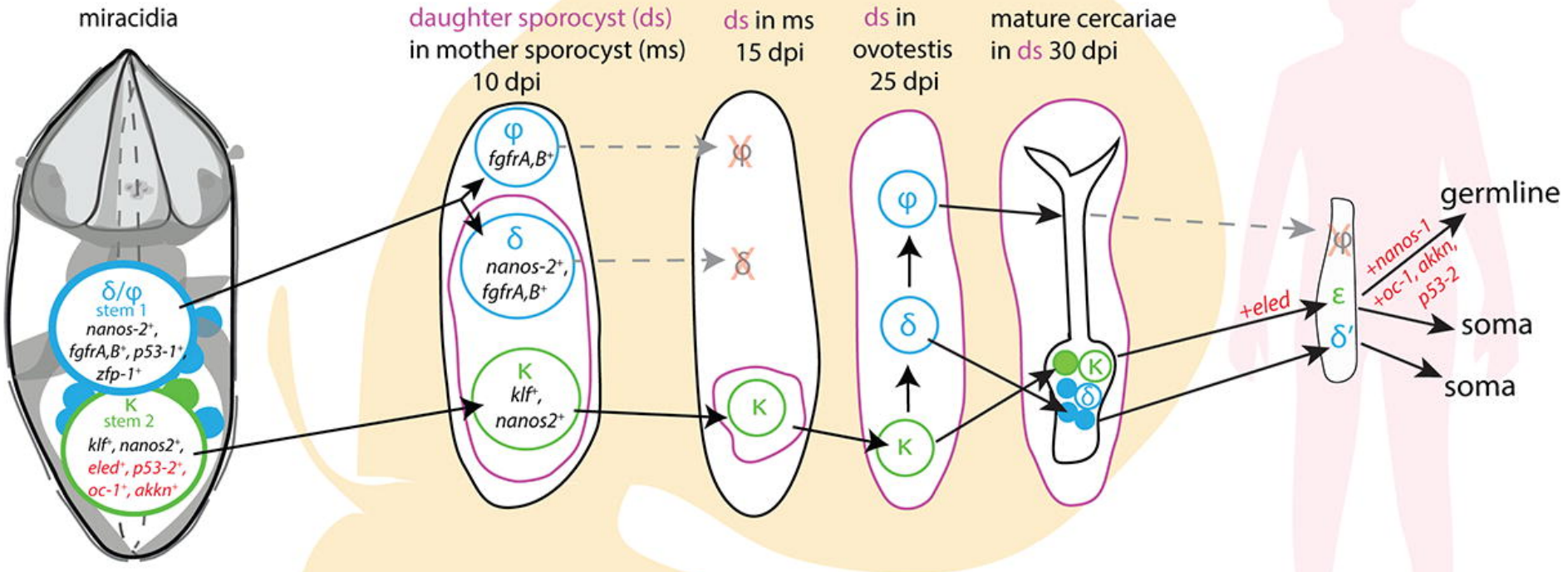

A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plate
Tegument
Protonephridia
Parenchyma 1
Parenchyma 2
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Myosin light chain 1 – 045220
Troponin t – 179810
Paramyosin (PRM) – 085540
Tropomyosin–1 – 340760
NA – 347710
TRPA – 342190
Collagen alpha chain – 159600
Titin – 126240
Kunitz-type protease inhibitor – 052230
Protein Wnt – 156540
Notum – 015820
PDZ domain protein – 311170
Ca-activated K channel – 156150
Myogenic factor MyoD – 167400
Ca-transporting ATPase – 136710
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expressed
0
25
50
75
Bi)
ii)
Pan muscle paramyosin (Smp_085540) DAPI Phalloidin
Ci)
ii)
Pan muscle paramyosin (Smp_085540) DAPI Phalloidin
Di)
ii)
Muscle 1 protease inhibitor (Smp_052230) Wnt (Smp_156540)
Ei)
ii)
iii)
Muscle 1 notum (Smp_015820) Wnt (Smp_156540)
Fi)
ii)
Muscle 2 Ca-activated K channel (Smp_156150) myoD (Smp_167400)


A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plates
Parenchyma 1
Parenchyma 2
Protonephridia
Tegument 1
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
Bi)
ii)
Pan neural cpx (Smp_050220)
DAPI
Ci)
ii)
Neuron 2 - Smp_201600
Neuron 3 -Smp_071050
Di)
ii)
iii)
Neuron 4 - Smp_319480
Ciliary plate - Smp_096390
Ei)
ii)
iii)
cpx
Neuron 5 - Smp_343710
F)
cpx
central
N 2
N 3
peripheral
N 4
N 5
ciliary plate
Gi)
I
II
III
IV
Ciliary plate -Smp_096390
ii)
iii)
Phalloidin


A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plates
Tegument
Protonephridia
Parenchyma 1
Parenchyma 2
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Percent Expressed
0
25
50
75
Dynein light chain – Smp_312630
Dynein light chain – Smp_302860
Dynein light chain – Smp_200180
Probable dynein light chain – Smp_201060
DUF3421 domain protein – Smp_331910
MEG 6 – Smp_163710
Putative nephrin – Smp_130070
Ovule protein – Smp_124290
Ovule protein – Smp_327570
Thymosin beta – Smp_025860
c-Myc-binding protein – Smp_030440
ShKT domain – Smp_335600
SPO-1 protein – Smp_341790
NA – Smp_337250
Inositol monophosphatase – Smp_323640
UPF0506 domain protein – Smp_331560
Zinc transporter Slc39a7 – Smp_318920
Zinc transporter Slc39a7 – Smp_318890
Antigen Sm21.7 – Smp_086480
DUF2955 domain protein – Smp_191970
Ribonuclease, putative – Smp_105100
V-type proton ATPase – Smp_054220
Bi)
ii)
DAPI
Tegument - Meg 6 (Smp_163710)
Protonephridia - Smp_335600
Parenchyma - Smp_318890
Ci)
ii)
Meg 6 (Smp_163710)
Phalloidin
DAPI
Di)
ii)
iii)
Meg 6 (Smp_163710)
Ciliary plate (Smp_096390)
Overlay
Ei)
ii)
AFC
ETN
PFC
DAPI
Phalloidin
Protonephridia ShKT domain Smp_335600
Fi)
ii)
iii)
Parenchyma Smp_318890
Pan stem ago2-1 (Smp_179320)
DAPI


A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Neuron 6
Parenchyma 1
Parenchyma 2
Protonephridia
Tegument
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
−1
NA – Smp_333010
14 kDa fatty acid–binding protein – Smp_095360
60S ribosomal protein L30 – Smp_044580
Uridine phosphorylase – Smp_308145
Nuclear transport factor 2–like protein – Smp_037700
Queuosine salvage protein – Smp_094920
NA – Smp_348820
Histone H3 – Smp_348800
Dnaj homolog subfamily B member 4 – Smp_020920
40S ribosomal protein S4 – Smp_011570
Dna–binding protein inhibitor – Smp_155360
60S ribosomal protein L21 – Smp_337330
Profilin splice variant 4 – Smp_080920
60S ribosomal protein L6 – Smp_340210
60S ribosomal protein L28 – Smp_084460
NA – Smp_201810
NA – Smp_202770
tRNA specific adenosine deaminase 2 – Smp_053780
NA – Smp_136950
NADH dehydrogenase subunit 6 (mt) – Smp_900020
Cytochrome c oxidase subunit 1 (mt) – Smp_900000
NADH dehydrogenase subunit 4 (mt) – Smp_900090
NADH dehydrogenase subunit 2 (mt) – Smp_900040
Tr–type G domain–containing protein – Smp_334080
NA – Smp_111340
YTH domain–containing protein – Smp_323830
Transcriptional cofactor CA150 – Smp_213300
leukocyte receptor cluster (Lrc) – Smp_097830
Putative unkempt protein – Smp_154250
ago1-2 – Smp_179320
Nanos–2 – Smp_051920
fgfrA – Smp_175590
fgfrB – Smp_157300
Kruppel–like factor – Smp_172480
P53– Smp_139530
Phosphomevalonate kinase – Smp_072460
Bi)
Protonephridia
Neuron 5
Stem D (Delta/Phi)
Stem B (Kappa)
Stem G
Neuron 6
Muscle 2
Neuron 1
Stem F
Stem E
Stem A (Kappa)
Neuron 2
Neuron 4
Tegument
Stem C (Delta/Phi)
Muscle 1
Parenchyma 1
Parenchyma 2
Neuron 3
UMAP_2
UMAP_1
ii)
Neuron 2
Parenchyma 2
Muscle 1
Neuron 3
Kappa (Stem A&B)
Parenchyma 1
Stem F
Neuron 1
Stem E
Stem G
Muscle 2
Neuron 6
Delta/Phi (Stem C&D)
Tegument
Neuron 4
Neuron 5
Protonephridia
UMAP_2
UMAP_1
Ci
ii
ago2-1 (Smp_179320)
Di
ii
p53 (Smp_139530) stem 1 Delta/Phi
UPPA (Smp_308140) stem 2 Kappa


A)
clustering of cells
Kappa-like
Stem A
Stem B
Stem G
Tegument
Stem
Stem F
Stem E
Tegument
Stem C
Stem D
sporocyst
miracidia
Delta/Phi-like
B)
1
2
3
4
5
6
0
C)
latent time
1
0
D)
Smp_139530 p53-1
velocity
expression
unspliced
spliced
E)
Smp-145470 zfp-1
velocity
expression
unspliced
spliced
F)
Smp-136160 p53-2
velocity
expression
unspliced
spliced
G)
Smp-041540 eled
velocity
expression
unspliced
spliced


miracidia
δ/φ
stem 1
nanos-2+,
fgfrA,B+, p53-1+,
zfp-1+
κ
stem 2
klf+, nanos2+,
eled+, p53-2+,
oc-1+, akkn+
daughter sporocyst (ds)
in mother sporocyst (ms)
10 dpi
φ
fgfrA,B+
δ
nanos-2+,
fgfrA,B+
κ
klf+,
nanos2+
ds in ms
15 dpi
ds in
ovotestis
25 dpi
mature cercariae
in ds 30 dpi
+eled
ε
δ′
+nanos-1
+oc-1, akkn,
p53-2
germline
soma
soma

Muscle (19%)
Nerve (57%)
Tegument (13%)
Protonephridia (2%)
Parenchyma (2%)
Stem (7%)