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: Kate Rawlinson and Matthew Berriman).

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 and *in-situ* hybridisation we identified every cell in the whole organism revealing the organisation of the miracidium. This single cell atlas provides the foundation to understand the development and interaction of cell types and tissues as they change over a life cycle that is characterised by complex morphological changes.

Introduction

The parasitic blood fluke *Schistosoma mansoni* (Phylum Platyhelminthes) 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 (skin) 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*., 2024). 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 (Nanes 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*., 2024), 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 specifically, 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 (parenchyma), stem/germinal cells, epithelial (ciliary plates and

101 epidermal ridges of the tegument cells) and terebratorium (an apical modified epidermal plate
102 with sensory structures and gland openings) (Pan, 1980).
103
104 To characterise this critical transmission stage, we have generated a single cell transcriptome
105 atlas for the miracidium. Our atlas provides molecular definitions for all described cell types
106 (Pan, 1980) and indicates the functions of different cell types within tissues. Multiplexed *in*
107 *situ* validation of cell type marker genes enabled the cell types to be spatially resolved within
108 the larva. Furthermore, the nuclear segmentation of the cells expressing tissue-specific genes
109 has revealed the relative contribution of each tissue type to the animal and each cell in the
110 larva has been assigned to a tissue. A key finding is the identification of two stem cell
111 populations in a miracidium, and advances in our understanding of the genome (Buddenborg
112 *et al*., 2021) has enabled us to determine that these two stem populations are transcriptionally
113 distinct between male and female larvae due to sex-linked gene expression. To predict the
114 fate of these two stem populations we carried out a trajectory analysis, combining scRNA-seq
115 data from miracidia and mother sporocysts, and this indicated that one stem cell population
116 likely contains the pluripotent cells that will develop into the cercariae, while the other
117 population gives rise to the somatic tissues, including the tegument. Similarities in gene
118 expression with the adult suggest that there is a tegument developmental program that is
119 redeployed at multiple stages of the complex life cycle.
120
121 **Results**
122 **Single-cell RNA-seq of 20,478 miracidia cells reveals 19 cell types.**
123 There are ~365 nuclei in a miracidium (median = 365, larva #1: 342, larva #2: 365, larva #3:
124 377). Some cell types, such as the apical gland and epithelial/tegument cells, are
125 multinucleate; and others, such as the ciliary plates may be anucleate (Pan, 1980, but cf
126 Meulemen *et al*., 1978). Therefore, the number of nuclei provides an approximation for the
127 actual number of cells (Figure 1A).
128
129 We performed scRNA-seq on a pool of ~20,000 mixed-sex miracidia, collected within four
130 hours of hatching from eggs (Figure 1B). Following dissociation, four samples of the cell
131 suspensions were collected; two were enriched for live cells using fluorescence-activated cell
132 sorting (FACS), and two were left unsorted to capture the maximum cellular diversity for the
133 atlas (Figure 1B). Using the droplet-based 10X Genomics Chromium platform,
134 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 (version 4.3.0) (Hao *et al*., 2021), 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). We used Seurat's JackStraw and ElbowPlot, along with molecular cross-validation to select the number of principal components, and Seurat's clustree to select a resolution where clusters were stable (Hao *et al*., 2021). Manually curated lists of previously defined cell-type-specific genes from later developmental stages (Diaz *et al*., 2020; Wendt *et al*., 2020; Nanes 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).

There were contributions from sorted and unsorted samples in almost all clusters (except ciliary plates). We found that some cell/tissue types had similar recovery from both methods (e.g. Stem A, Muscle 2, and Tegument), others were preferentially recovered by sorting (e.g Neuron 1, Neuron 4, and Stem E), and some were depleted by sorting (e.g. Parenchyma 1, Protonephridia, and Ciliary plates) (Supplementary Figure 1), Supplementary Table 4). This variation in recovery, therefore, enabled us to maximise the discovery and inclusion of different cell types in the atlas.

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*., 2024), 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

169 conserved transcriptional regulator of muscle (Tarashansky *et al*., 2021) was also expressed
170 in both muscle clusters. High expression of *paramyosin* (*PRM*; Smp_085540) was seen in
171 muscle cell clusters 1 and 2 (Figure 2A). *In situ* hybridisation (ISH) (Figure 2B & C;
172 Supplementary Video 1) and image analysis (Supplementary Video 2) identified 74 nuclei
173 surrounded by *PRM* transcripts (i.e. ~19% of cells in a miracidium are muscle cells).
174 Counterstaining with phalloidin (to label F-actin filaments) revealed an orthogonal grid of
175 circular and longitudinal body wall muscles and from this we identified which *PRM*$^{+}$ nuclei
176 belonged to each muscle subtype. The nuclei of the circular muscles formed two distinct
177 bilaterally symmetrical lines that ran peripherally from pole to pole of the larva (~28 nuclei in
178 total). Thirty-three more *PRM*$^{+}$ nuclei sat regularly spaced between the circular muscle nuclei
179 and corresponded to longitudinal muscles. Another 13 *PRM*$^{+}$ nuclei formed a unilateral
180 cluster adjacent to the apical gland cell (identifiable by its four nuclei; Pan, 1980) (Figure
181 2Cii; Supplementary Video 1 and 2).
182
183 Muscle cluster 1 (2,407 cells) was distinguished from muscle cluster 2 by markedly higher
184 expression of a Kunitz-type protease inhibitor (Smp_052230; Figure 2A), and ISH showed
185 Smp_052230 transcripts along the two peripheral lines of circular muscle nuclei (Figure 2D).
186 Subpopulations of muscle cluster 1 also expressed orthologs of *Wnt-11-1* (Smp_156540) and
187 *Notum* (Smp_015820), known anterior-posterior body axis patterning genes, where Notum
188 acts as an inhibitor of Wnt signalling (Peterson and Reddien, 2009; Kakugawa *et al.,*
189 2015)(Figure 2A). *Notum* (Smp_015820) was highly expressed in the cluster of 13 muscle
190 nuclei adjacent to the apical gland cell (Figure 2E). *Wnt-11-1* (Smp_156540) was expressed
191 at the posterior pole, in seven of the Kunitz-type protease inhibitor-expressing circular muscle
192 cells (Figure 2D; Supplementary Video 3). *Notum* was also expressed, albeit weakly, in some
193 nuclei of circular muscles, including those expressing *Wnt-11-1* (Supplementary Video 4).
194
195 Muscle cluster 2 (1,355 cells) was distinguished by higher expression of genes encoding a
196 PDZ domain protein (Smp_311170) and *Calcium-transporting ATPase* (Smp_136710), as
197 well as the unique expression of a putative calcium-activated potassium channel
198 (Smp_156150), and the transcription factor *MyoD* (Smp_167400)(Figure 2A); ISH of the
199 latter two genes showed expression only in the longitudinal body wall muscles (Figure 2F;
200 Supplementary Video 5). Orthologs of other axial patterning genes, *Netrin receptor UNC5*
201 (Smp_172470), *Wnt5* (Smp_145140), and a putative wnt inhibitor frizzled, frzb2
202 (Smp_062560)(Witchley *et al*., 2013), showed distinct expression in Muscle 2 cells.

203

204 The higher expression of Muscle 1 markers in circular body wall muscles (BWM) and
205 Muscle 2 genes in the longitudinal BWM revealed distinct transcriptomic signatures for these
206 two types of myocyte that make up the orthogonal body wall musculature. In fact, differential
207 expression analysis showed that *Wnt-11-1* (circular) and *MyoD* (longitudinal) were among
208 the most differentially expressed genes between the two muscle clusters (Supplementary
209 Table 5).

210 **Neural abundance and diversity**

211 We uncovered five clusters expressing the neural markers *Complexin* (*cpx*; Smp_050220),
212 *prohormone convertase 2* (*PC2*; Smp_077980) and Smp_068500 (a neural marker in adult
213 worms, Wendt *et al*., 2020) (Figure 3A). *Complexin* was expressed in and around at least 209
214 nuclei, indicating that 57% of cells in the miracidium were neurons; out of which 129 *cpx*$^{+}$
215 nuclei formed the nuclear rind of the brain (or neural mass/ring), and the remaining 80 were
216 situated peripherally, either anterior or posterior to the brain (Figure 3B). Sensory organelles
217 were identified for peripheral neural cell types. We were able to validate neuron clusters 2 - 5
218 using ISH of specific marker genes, and performed an *in silico* analysis of Neuron 1.

219

220 Neuron cluster 1 was large (1,470 cells), and expressed neuron-marker *complexin*, the
221 neuroendocrine precursor *7B2* (Smp_073270) and an uncharacterised gene (Smp_068500).
222 At this resolution, there were no clear marker genes restricted to Neuron 1. We explored
223 whether Neuron 1 could be further subdivided into transcriptionally distinct cells by
224 subclustering (Supplementary Figure 2; Supplementary Table 6) using the self-assembling
225 manifold (SAM) algorithm (Tarashansky *et al*., 2019) with ScanPy (Wolf *et al*., 2018), given
226 its reported strength in discerning subtle variation in gene expression (Tarashansky *et al*.,
227 2019), although a similar topology was subsequently found using Seurat. Three subclusters of
228 Neuron 1 expressed neuropeptide precursor genes amongst their top five markers; *PWamide*
229 (Smp_340860)(Wang *et al*., 2016) in subcluster 14, *Neuropeptide F* (Smp_088360) in
230 subcluster 8, and *Neuropeptide Y* (Smp_159950) in subcluster 13 (Supplementary Figure 2);
231 and each was expressed in the nuclear rind of the brain (Supplementary Figure 3). Another
232 subcluster was composed of putative photoreceptor cells (subcluster 7 - Supplementary
233 Figure 2, Supplementary Table 6) because their top-ten markers include two opsins
234 (Smp_180350 and Smp_332250) – identified here as a rhabdomeric opsin and a peropsin,
235 respectively (Supplementary Figure 4). Subcluster 7 also expressed other conserved
236 phototransduction cascade genes, such as a 1-phosphatidylinositol 4,5-bisphosphate

237 phosphodiesterase (Smp_245900) and a beta-arrestin (Smp_152280) (Supplementary Figure
238 2, Supplementary Table 6).
239
240 Neuron 2 expressed two further rhabdomeric opsins (Smp_180030 and Smp_104210;
241 Supplementary Figure 4), but across the cluster of 282 cells, the most expressed marker was
242 the uncharacterised gene Smp_201600. ISH revealed three strongly Smp_201600^{+} cells, with
243 nuclei in a median posterior location of the brain, and three cells with weaker expression in a
244 more anterior position (Figure 3C, Supplementary Video 6). Other top markers include *7B2,*
245 a guanylate kinase (Smp_137940) and a putative homeobox protein (Smp_126560).
246
247 Neuron 3 was a small cluster (47 cells) expressing *complexin* and *7B2*. The uncharacterised
248 gene, Smp_071050, was unique to and highly expressed throughout this population (Figure
249 3A). ISH showed expression in ~6 cells with projections that extended into the brain and to
250 the periphery, three of which sat adjacent to three Neuron 2 cells. There was also weak
251 expression in the four nuclei of the apical gland cell (Figure 3Ci, Supplementary Video 6).
252 STRING analysis of the top 100 markers of Neuron 3 predicted two protein interaction
253 networks with functional enrichment; these were associated with 'synapses' and 'calcium'
254 (Supplementary Figure 6). In Neuron clusters 1, 2 and 3, there was an enrichment for genes
255 annotated with Gene Ontology (GO) terms related to 'neuropeptide signalling pathway' and
256 'neurotransmitter transport' (Supplementary Table 7). There were, therefore, at least 5, and
257 likely several more, transcriptionally distinct cell types in the brain mass (we validated three
258 subclusters of Neuron 1, as well as Neurons 2 & 3).
259
260 Cells from neuron clusters 4 and 5 had nuclei peripheral to the brain. Neuron 4 formed a
261 large (1,140 cells), discrete cluster. The top marker was an uncharacterised gene,
262 Smp_319480 (Figure 3A), that, based on protein-structure prediction using I-TASSER (Yang
263 & Zhang, 2015), was similar to human mTORC1 (PDB entry 5H64; TM-score 0.85). *In situ*
264 expression revealed ~26 Smp_319480^{+} cells; four at the anterior end, with projections
265 leading to the body wall in the first tier of ciliary plates, and ~22 sat lateral and directly
266 posterior to the brain. These latter cells formed four clusters of ~3–7 cells, with Smp_319480
267 transcripts distributed around the nuclei and along projections leading into the brain and
268 extending out to the body wall terminating between the second and third tier of ciliary plates
269 (Figure 3D, Supplementary Video 9 and 10). Twenty-three multi-ciliated nerve endings
270 formed a girdle around the larva between the second and third tier of ciliary plates

271 (Supplementary Figure 3C), and Smp_319480 transcripts were expressed at the base of these
272 cilia (Supplementary Figure 3D); some transcripts also reached the nuclei of circular body
273 wall muscles (Supplementary Figure 3E), suggesting these cells may function as sensory-
274 motor neurons. Two genes, identified as top markers and co-expressed in many Neuron 4
275 cells, encode likely orthologues of *Polycystin 1* (Smp_165230) and *Polycystin 2*
276 (Smp_165660) (Supplementary Figure 5A), with the latter prediction strongly supported by a
277 Foldseek search (van Kempen *et al*., 2023) against the protein database (PDB). Polycystins 1
278 and 2 form a heterodimeric ion channel required for oscillating calcium concentrations within
279 vertebrate cilia (Kim *et al*., 2016). A Sox transcription factor, Smp_301790, was also
280 expressed in Neuron 4 and co-expressed with *Polycystin 1* and *2* (Supplementary Figure 5B).
281 Other top markers included a putative potassium channel (Smp_141570), a cAMP-dependent
282 protein kinase (Smp_079010), a GTP-binding protein (Smp_045990) and an ortholog of
283 human *fibrocystin-L* (Smp_303980). The GO terms 'cilium', 'regulation of apoptotic
284 process' and 'potassium ion transmembrane transport' were significantly enriched, and
285 STRING analysis showed networks enriched in 'plasma membrane-bound cell projection'
286 (Supplementary Table 7 Supplementary Figure 6).
287
288 Neuron 5 cluster comprised 276 cells. An EF-hand domain-containing protein, Smp_343710,
289 was highly expressed in 55% of these cells. ISH showed expression in 10-20 cells whose
290 nuclei sat outside of, and anterior to, the brain, in and around the paired lateral glands and
291 their secretory ducts (Supplementary Video 11). There was also expression in a pair of
292 bilaterally symmetrical bulbous protrusions (called lateral sensory papilla and thought to be
293 depth sensors, Pan, 1980)(Figure 3F). STRING analysis revealed networks associated with
294 'synapses', 'ion channel binding' and 'cation transmembrane transporter activity'
295 (Supplementary Figure 6).
296
297 We expected to find a discrete cluster(s) for the penetration glands, and that it would show
298 similarities to the neural clusters (as glandular cells arise from neuroglandular precursor cells
299 in other animals, such as the sea anemone, *Nematostella vectensis*, Steger *et al*., 2022). There
300 are three gland cells per miracidium - one apical and two lateral (Pan, 1980). While the top
301 marker gene for Neuron 3 was expressed in the four nuclei of the apical gland (Figure 3Ci),
302 and there was expression of a Neuron 5 marker in the lateral glands (Figure 3Fi), these genes
303 were expressed in many other cells as well. This suggests that gland cells are either sub-
304 clusters of Neurons 3 and 5 or they were not captured in the 10X GEMs. To characterise the

305 transcriptomes of the gland cells, we carried out plate-based single cell RNA-seq (New
306 England BioLabs) on 37 manually selected gland cells (Supplementary Table 8). The
307 biological process GO terms enriched in the most abundant genes (top 200 genes by median
308 transcripts per million, Supplementary Table 8) included translation, protein folding and ATP
309 synthesis coupled electron transport (Supplementary Table 9). The gland cells showed co-
310 expression of neural markers (*Complexin* and *7B2*) and muscle markers *Myosin light chain 1*
311 (Smp_045200), *Troponin t* (Smp_179810) and *Actin-2* (Smp_307020) (Supplementary
312 Figure 7). To identify genes that were specific to gland cells, we searched for differentially
313 expressed genes in the gland cells compared to the other cells in the plate-based scRNA-seq
314 data (i.e. ciliary plates and unknown cells) (Supplementary Table 10 and 11), and then looked
315 at their expression in the 10X-based scRNA-seq data. A lack of, or minimal, expression of
316 these markers in the 10X-based scRNA-seq data, would support the hypothesis that they were
317 specific to the gland cells. Five venom antigen-like (VAL) genes (belonging to the
318 SCP/TAPS or CAP) superfamily) were expressed in the gland cells; Smp_120670 (VAL 5)
319 and Smp_070250 (VAL 15) in over 60% of the gland cells, and Smp_176180 (VAL 9),
320 Smp_002630 (VAL 2) and Smp_331830 (SCP protein) in less than 50%. Many gland cells
321 expressed multiple VAL genes (Supplementary Figure 7). In the 10X-based scRNA-seq data,
322 very few cells expressed these VAL genes (1-5 cells for each gene, distributed between
323 Neurons 1 and 5, Muscle 1, Parenchyma 2, and Stem F and G clusters). The VAL genes
324 expressed in the gland cells are uniquely expressed during the egg, miracidia and sporocyst
325 stages (Yoshino *et al*., 2014; Lu *et al*., 2018; Buddenborg *et al*., 2023), and it has been
326 reported that VAL 9 (Smp_176180) induces the differential expression of an extracellular
327 matrix remodelling gene in the snail host (Yoshino *et al*., 2014). Other differentially
328 expressed genes in the gland cells included Smp_303400 (an HSP70 homolog), Smp_179250
329 (*α-galactosidase/α-N-acetylgalactominidase*), Smp_317530 (uncharacterised protein), and
330 two putative major egg antigens Smp_185680 and Smp_303690 (Supplementary Figure 7).
331 Eight putative secreted proteins previously identified in miracidia and *in vitro* transformed
332 miracidia (Wang *et al*., 2016; Wu *et al*., 2006) were expressed in the glands cells (and not in
333 other cell types based on the 10X data), and five of these showed relatively high expression;
334 Smp_302170 (*HSP70* homolog), Smp_303400 (*HSP70* homolog), Smp_049230 (*HSP20*),
335 Smp_304250 (*purine nucleoside phosphorylase*), and Smp_302180 (putative heat shock
336 protein 70)(Supplementary Figure 7). Typical of venom glands in other animals, the
337 transcriptomes of the miracidia gland cells are enriched in genes involved in protein

338 translation, stabilisation and stress response to cope with mass protein production (Zancolli *et*
339 *al*., 2021).

340 **Ciliary plate gene marker identifies six submuscular cell bodies**

341 A small cluster of 40 cells (almost exclusively from the unsorted samples) was identified near
342 the tegument cluster (Figure 1C). The top markers were Smp_096390, a 16 kDa calcium-
343 binding protein (CaBP) and *p25 alpha* (Smp_097490, a marker for ciliated neurons in adult
344 worms, Wendt *et al*., 2020) (Figure 3A). *In situ* hybridisation revealed CaBP and *p25 alpha*
345 expression in the ciliary plates; epithelial cells with thousands of motile cilia used for
346 swimming (Figure 3G and Supplementary Figure 3). Both genes were also expressed in six
347 previously undescribed cells that sent a projection out to the exterior (between the first and
348 second tier of ciliary plates as seen with phalloidin staining, Figure 3G) and whose nuclei sat
349 below the body wall musculature (Supplementary Videos 7 & 8; Supplementary Figure 3F).
350 Given the association of *p25 alpha* with ciliated neurons in adults (Wendt *et al.*, 2020),
351 expression in the sub-muscular cells suggests they project cilia into the external environment
352 and may be sensory cells. The expression of CaBP and *p25* in both the ciliary plates and
353 submuscular cells suggests that these anatomical structures are functionally linked. GO terms
354 for this cluster were related to 'mitochondrial transmembrane transport' and 'microtubule-
355 based process' (Supplementary Table 7), and STRING analysis showed two interaction
356 networks: one associated with 'carbon metabolism and mitochondria envelope', the second
357 associated with 'cytoplasmic dynein complex' and 'dynein light chain superfamily'
358 (Supplementary Figure 6). Further plate-based scRNA-seq of 21 individual ciliary plates
359 confirmed the high expression of CaBP (Smp_096390) and *p25 alpha* (Smp_097490), with
360 more than two-fold higher expression of CaBP than any other gene by median TPM
361 (transcripts per million) (Supplementary Table 12). Although the ciliary plates expressed *p25*
362 (a ciliated neuron marker in the adult, Wendt *et al.,* 2020), they did not express many of the
363 typical neural markers, e.g. *complexin* and *7B2* (Supplementary Figure 7).
364

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

367 We identified a cluster of 366 cells as tegumental based on the expression of markers for
368 tegument (skin) in the adult worm (Wendt *et al.*, 2020), e.g. genes encoding *dynein light*
369 *chain* (Smp_302860, Smp_200180), a DUF3421 domain-containing protein (Smp_331910)

370 and micro exon gene *MEG6* (Smp_163710). Two other dynein light chain genes were top
371 markers (Smp_201060, Smp_312630), and both a tetraspanin (Smp_303290) and *Gelsolin*
372 (Smp_008660) were specific to this cluster. *Meg6* (Smp_163710) was expressed in 66% of
373 the cells in this cluster; ISH showed 46 *Meg6*$^{+}$ cells, and the nuclei were in the posterior two-
374 thirds of the larva. The nuclei sat below the body wall muscle, and cytoplasmic protrusions
375 reached between muscle filaments and formed the epidermal ridges between the ciliary plates
376 (Figures 4C & D). These cells correspond to the epidermal ridge cells described by Pan
377 (1980). A second tegument marker, a dynein light chain (Smp_200180), showed overlapping
378 expression with *Meg6* (Supplementary Video 15). Enriched GO terms for tegument genes
379 included 'microtubule-based process' and 'protein folding' (Supplementary Table 7).
380 STRING analysis revealed a network of 25 genes associated with protein localisation to the
381 ER, folding, export and post-translational protein targeting to the membrane (Supplementary
382 Figure 8).
383
384 **Spatial localisation of top marker transcripts identifies a simple protonephridial system**
385 There were 536 cells in the protonephridial cluster, expressing genes common to excretory
386 systems in other animals (Gąsiorowski *et al.,* 2021), e.g., the structural protein *nephrin*
387 (Smp_130070) and the transcription factor *odd-skipped-related* (Smp_152210). Eighty
388 percent of the cells expressed the protonephridia marker Smp_335600 (ShKT-domain
389 containing protein) (Nanes Sarfati *et al.*, 2021) (Figure 4A). Smp_335600 was expressed
390 around six nuclei in a bilaterally symmetrical pattern (Figure 4Ei) (Supplementary Video 12).
391 On each side, Smp_335600 transcripts were distributed along the s-shaped path of an
392 excretory tubule, passing its nucleus, and connecting the anterior and posterior flame cells
393 (Figure 4Ei). Within the flame cells, expression was detected around the nuclei rather than
394 the barrel (Figure 4Eii). The four ciliated flame cells and two excretory ducts comprise the
395 excretory system of the miracidium. Accordingly, STRING (Supplementary Figure 8) and
396 GO analyses (Supplementary Table 7) of this cluster confirmed enrichment for terms
397 including 'cilium organisation', 'axoneme assembly', 'cytoskeleton organisation' and
398 'microtubule-based process'.
399
400 **Two distinct parenchyma clusters predicted in the miracidium**
401 We identified two clusters of cells (1,093 cells from Parenchyma 1 and 605 cells from
402 Parenchyma 2) that were transcriptionally similar to the parenchymal cells found in later
403 developmental stages; e.g. they express hypothetical protein (Smp_063330), *cathepsin B-like*

404 *cysteine proteinase* (Smp_141610) and *serpin* (Smp_090080) (Diaz *et al.,* 2020; Wendt *et al.*,
405 2020; Diaz & Attenborough *et al*., 2024). All top marker genes for Parenchyma 1 were also
406 expressed in Parenchyma 2 but at lower levels, whereas the DUF2955 domain-containing
407 protein (Smp_191970) was uniquely expressed in Parenchyma 2 (Figure 4A). A gene
408 encoding a Zinc transporter Slc39a7 (Smp_318890) was a top marker for both clusters, with
409 expression confirmed *in situ* within 7–9 cells; two of which were located anteriorly on either
410 side of the brain and 5-7 located posterior to the brain (Figures 4B & F). These parenchyma
411 cells have long cytoplasmic protrusions that reach between all the other cells, filling the
412 intercellular space (Figure 4Fii-iii; Supplementary Videos 13 & 14). Pan (1980) called these
413 the interstitial cells and, observing abundant glycogen particles and lipid droplets, proposed
414 they function as sources of energy for the non-feeding larva and as nurse cells for the stem
415 cells during the first few days post-infection of the snail. We performed ISH for two
416 Parenchyma 2 markers, Smp_191970 and Smp_320330, but the first showed no detectable
417 signal, and the second had broad expression across all cells with no specific overlap with the
418 pan-parenchymal marker Smp_318890. Therefore, by ISH, we could only identify
419 parenchymal cells in general and could not identify the two clusters *in situ*. STRING analysis
420 for the Parenchyma 1 showed enrichment in networks for 'carbohydrate metabolic process'
421 and 'hydrolase', and for Parenchyma 2: 'phagosome', 'mTOR signalling pathway',
422 'mitochondrial inner membrane' and 'proton-transporting two sector atpase complex'
423 (Supplementary Figure 9). Analysis of GO annotations was able to distinguish between the
424 two parenchyma sub-populations, with Parenchyma 1 genes enriched for 'extracellular
425 region' and 'endopeptidase inhibitor activity', and Parenchyma 2 genes enriched for
426 processes that are 'dynein complex' and 'microtubule-based process' (Supplementary Table
427 7; Supplementary Figure 10).
428
429 **Two major stem cell populations in the miracidium that clustered by sex**
430 We identified stem cells based on the expression of known stem cell markers for the mother
431 sporocyst stage, e.g. *ago2-1* (Smp_179320), *nanos-2* (Smp_051920), *fgfrA* (Smp_175590),
432 *fgfrB* (Smp_157300), *klf* (Smp_172480) and *hesl* (Smp_024860)(Wang *et al*., 2013;
433 2018)(Figure 5A). The mother sporocyst has been reported to have three stem cell
434 populations: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*fgfrA*$^+$, *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$,
435 *fgfrB*$^+$ and *hesl*$^+$) based on scRNAseq of 35 stem cells (Wang *et al*., 2018). Wang *et al*. (2018)
436 suggest that the Kappa population serve as 'embryonic' stem cells that give rise to both the
437 Delta and Phi populations; Delta cells generate somatic tissues and Phi produce temporary

438 larval tissues, including tegument. In the miracidium, we identified seven stem cell clusters
439 by expression of *ago2-1* (10,686 cells): Stem A and B are *klf*$^{+}$ and *nanos-2*$^{+}$ (i.e. Kappa-like);
440 Stem C and D are *nanos-2*$^{+}$, *fgfrA*$^{+}$, *fgfrB*$^{+}$, and *hesl*$^{+}$ (i.e. they resembled both Delta and
441 Phi)(Figure 5A, Supplementary Figure 11); as well as Stem E, F and G (Figure 5B).
442
443 Working with the latest version of *S. mansoni* genome (v10), where the sex chromosomes Z
444 and W have been resolved (Buddenborg *et al*., 2021), we were able to determine that the
445 Delta/Phi-like and Kappa-like populations are each composed of two well-separated clusters
446 that can be distinguished based on the expression of sex-linked genes (Figure 5A;
447 Supplementary Figure 12). *Schistosoma mansoni* has 7 pairs of autosomes and one pair of sex
448 chromosomes; males are homogametic (ZZ) and females heterogametic (ZW) (Buddenborg
449 et al., 2021). The sex chromosomes are composed of sex-specific regions that are unique to
450 each chromosome flanked by pseudoautosomal regions that are common to both Z and W. To
451 unveil the genes contributing to the sex-driven clustering we removed in turn different sex
452 chromosome regions from the analysis. When all Z-specific region (ZSR) and W-specific
453 region (WSR) genes were removed from the dataset and the cells reclustered, the two
454 Delta/Phi-like and two Kappa-like clusters each collapsed into single stem cell clusters,
455 giving five stem cell clusters in total (i.e. Delta/Phi, Kappa, E, F and G; Figure 5B,
456 Supplementary Figure 13). Removing from the dataset genes located solely on the WSR ($n \approx$
457 35), the ZSR ($n \approx 900$), or just the ~35 pairs of genes with homologous copies on the WSR
458 and ZSR did not impact the overall UMAP topology, i.e. in all three scenarios the stem cell
459 clusters remained separated by sex (Supplementary Figure 12). ZSR or WSR expression is
460 therefore sufficient on their own to split each of these stem clusters in two (Supplementary
461 Figure 12).
462
463 The unexpected subclustering of Delta/Phi and Kappa stem cells by sex could be either due to
464 the contribution to each cell's transcriptome by WSR genes (in females) that have diverged
465 from their ZSR homologues, or by differences in expression of ZSR genes between male or
466 female cells (see GO analysis of ZSR and WSR genes, Supplementary Table 13). Given the
467 small number of WSR genes, differences in ZSR genes — caused by a difference in copy-
468 number — seems likely to play a larger part. In many animals, dosage compensation
469 equalises gene expression across the sex chromosomes that are unevenly distributed between
470 the sexes. To investigate whether there is an incomplete or a lack of dosage compensation in
471 the ZSR genes in the *S. mansoni* miracidium, we classified all cells based on the expression

472 of WSR genes (i.e. only found in female individuals), so each cell was designated as W^+
473 (likely female) or W^- (likely male) (Supplementary Figure 12 A&B, Supplementary Figure
474 14A&B). The difference in ZSR and WSR gene expression was most pronounced in the
475 Delta/Phi-like and Kappa-like cells, shown by the separation of W^+ and W^- cells
476 (Supplementary Figure 14B). However, this analysis revealed other tissue-specific sex
477 differences in ZSR gene expression e.g. in the protonephridia, which was evident through
478 segregation of W^+ and W^- cells within that cluster. We found that W^+ and W^- cells were well
479 mixed in the majority of the remaining clusters (Supplementary Figure 14B). Male cells have
480 two copies of the ZSR (ZZ) and produce higher coverage across genes in this region
481 compared to female cells, which have one ZSR and one WSR (ZW). These findings suggest
482 that sex-linked gene dosage is not fully compensated in the miracidium, and there may be
483 tissue-specific variation.
484
485 To identify the stem cells *in situ*, we investigated the expression of the pan-stem marker
486 *ago2-1*. *Ago2-1*$^+$ cells were distributed lateral and posterior to the brain (Figure 5Ci) and
487 there was variation in the number between larvae, ranging from 20-29 (segmentation of
488 *Ago2-1*$^+$ nuclei in 5 larvae, median = 23, mean = 23.4)(i.e. ~7% of cells that make up the
489 larva are stem cells). Our count is slightly higher than previous estimates of 10-20
490 stem/germinal cells (Pan, 1980; Wang *et al*., 2013). Cells in the Delta/Phi-like cluster
491 expressed a canonical *p53 tumour suppressor family gene* (*p53-1*, Smp_139530) as a top
492 marker; *in situ* expression showed 12–14 *p53-1*$^+$ cells (segmentation of 3 larvae)(Figure 5D).
493 Cells in the Kappa-like cluster expressed *Uridine phosphorylase A* (*UPPA*, Smp_308145) as
494 a top marker; *in situ* expression showed 8–9 *UPPA*$^+$ cells, and they were located amongst the
495 *p53-1*$^+$ cells (segmentation of 3 larvae)(Figure 5D). We were not able to detect cells from the
496 Stem E, F and G clusters in the miracidium using ISH as there were no predicted marker
497 genes that were unique to these clusters. Although *Phosphomevalonate kinase* (*pmk*)
498 (Smp_072460) was a predicted marker for the Stem E cluster (Figure 5A), when multiplexed
499 as *p53/UPPA/pmk*, no *pmk*$^+$-only cells could be found (Supplementary Video 16). Likewise,
500 multiplexing *ago2-1/ p53/ UPPA* revealed no *ago2-1*$^+$-only cells (Supplementary Video 17).
501 Overall, we could spatially validate two stem cell populations within the miracidia: Delta/Phi
502 and Kappa. Cells in Stem E, F and G *in silico* clusters might be stressed/damaged/dying cells
503 or cells in transcriptionally transitional states (e.g. some *complexin* expression was also found
504 in a sub-population of Stem F).
505

506 A lack of EdU incorporation into miracidia cells (Supplementary Figure 15) indicated that
507 DNA replication does not take place in the stem or somatic cells during this free-swimming
508 stage. As the first stem cells start proliferating in the mother sporocyst ~20 hours post-
509 infection (Wang *et al.,* 2013), this suggests that both stem cell populations pause cell division
510 for at least 26 hours.
511
512 Many top markers for the Delta/Phi and Kappa stem clusters were ribosomal proteins (Figure
513 5A). STRING analysis showed that most of the top 100 markers of each cluster both formed
514 large single networks enriched for genes involved in translation, ribosome biogenesis and
515 gene expression, and this was particularly striking in the Delta/Phi cluster (Supplementary
516 Figure 16). Amongst the top genes differentially expressed between Kappa and Delta/Phi
517 (Supplementary Figure 17A; Supplementary Table 14) were six transcription factors up-
518 regulated in Delta/Phi: an ortholog of *lin-39* (Smp_247590), *tiptop* (Smp_180690), *hes-*
519 *related* (Smp_024860), *p53-1* (Smp_139530), *Zfp-1* (Smp_145470) and a putative
520 *homothorax homeobox protein/six-3-1* (Smp_147790). The simultaneous co-expression of
521 multiple transcription factors, indicates a multipotent progenitor stage (Kumar *et al*., 2019).
522 Enriched in, or specific to, Kappa cells were genes involved in glycolysis and lipid
523 metabolism, i.e. malate dehydrogenase (Smp_129610), 14 kDa fatty acid-binding protein
524 (Smp_095360) and a very long chain fatty acid elongase (Smp_051810). GO analysis of
525 differentially expressed genes (Supplementary Table 15) revealed that the most significantly
526 upregulated genes in Delta/Phi were related to the structural component of the ribosome and
527 translation; these are the first markers of a pre-activation stage in stem cells (Shin *et al*.,
528 2015). Whereas GO analysis showed that up-regulated genes in Kappa are involved in
529 transcription and metabolism (Supplementary Figure 17B), which indicated quiescent rather
530 than primed and activated stem cells (Llorens-Babadilla *et al*., 2015).
531
532 **One miracidia stem population contributes to somatic tissues and the other to the**
533 **germline.**
534
535 The Delta/Phi stem population simultaneously co-expressed three transcription factors that
536 are known to be specific to the flatworm tegument/epidermal lineage, i.e *p53-1, zfp-1, six3-1*
537 (Cheng *et al*., 2018; Collins *et al*., 2016; Wendt *et al*., 2018; 2022). This suggests that the
538 Delta/Phi stem population may have been the origin of the miracidia tegument cells and may
539 be the source of new tegument cells in the mother sporocyst stage. To explore the fates of the

540 two miracidia stem populations, we investigated the dynamics of their gene expression in
541 stem and tegument cells across the transition from miracidia to sporocyst. To do this, we
542 combined single-cell data from these developmental stages and cell types (this study and
543 Diaz & Attenborough *et al*., 2024) and used RNA velocity analysis (Bergen *et al*., 2020) to
544 estimate relative induction/repression and timing of gene expression in this dataset.

545 A UMAP plot of the combined clustering shows the gene expression dynamics of the
546 miracidia and sporocyst tegument and stem cells (Figure 6) (batch-corrected samples showed
547 similar results, Supplementary Figure 18). The velocity field and latent time representations
548 indicate the progression of some miracidia Delta/Phi stem cells towards the miracidia
549 tegument cells and a high speed of differentiation was inferred from the miracidia tegument
550 cells to the sporocyst tegument (Figure 6B & C). This suggests that the miracidia tegument
551 cluster may contain cells that range from post-mitotic stem cell progeny to tegument
552 progenitors to terminally differentiated tegumental cells.

553 Investigations into tegument development in adult schistosomes has revealed key regulators
554 and genes associated with different development states (Collins *et al*., 2016; Wendt *et al*.,
555 2018; Wendt *et al*., 2022). We examined the dynamics of these genes in the combined
556 miracidia/mother sporocyst dataset using RNA velocity and found that the velocities of these
557 genes show different temporal dynamics. *P53-1* (Smp_139530) was induced in the Delta/Phi
558 stem cells but repressed in some miracidia tegument cells and expression decreased in a
559 continuum across the miracidia tegument cells (Figure 6D). The flatworm-specific
560 transcription factor *Zfp-1* (Smp_145470) showed a similar expression pattern to *p53-1*,
561 except that the predominance of spliced copies (Figure 6E) suggested that transcription
562 happened earlier, during embryogenesis. A second zinc finger protein, *Zfp-1-1*
563 (Smp_049580) was expressed in tegumental cells more than stem cells, which suggested a
564 role in stem cell progeny and tegument progenitors (Supplementary Figure 19). In adults,
565 tegument cells are renewed continuously by a population of *tsp2*+ (Smp_335630) progenitor
566 cells that fuse with the tegument (Wendt *et al*., 2018). In the miracidia, *tsp2* was expressed in
567 the Delta/Phi stem cells as well as tegument cells (Supplementary Figure 19), suggesting
568 earlier expression than in the adult, or that Delta/Phi are progenitor cells. Several markers of
569 mature tegument cells in the adult were expressed in the miracidia and sporocyst tegument
570 cells but not in the stem cells, e.g. *annexin* (Smp_077720), *calpain* (Smp_241190) and
571 *endophilin B* (Smp_003230) (Supplementary Figure 19). *Annexin* is specific to the tegument

572 cells, and the phase plot showed the accumulation of mature transcripts in the sporocyst
573 tegument, while there was evidence of recent transcription in the miracidia tegument
574 (Supplementary Figure 19). A putative ortholog of a planaria transcription factor *six-3-1*
575 (Smp_147790), a marker for early epidermal progenitors (Cheng *et al*., 2018) was expressed
576 in the Delta/Phi stem cells and miracidia tegument (Supplementary Figure 19). All of these
577 genes showed little expression in the Kappa stem population (Supplementary Figure 19). The
578 expression of these known tegument/epidermal development and lineage-specific genes in
579 the Delta/Phi stem cluster, and the indication that *zfp-1* transcription occurred pre-hatching,
580 suggested that this stem population was the origin of the miracidia tegument cells and is the
581 source of the tegumental lineage. This, together with the observation that more than 80% of
582 protein-coding genes in the genome are detected in these miracidia data, suggests that similar
583 molecular programs for the development of tissues in the miracidia and mother sporocyst
584 may be redeployed multiple times across the life cycle.

585

586 As well as allowing us to interrogate the dynamics of known marker genes (Supplementary
587 Figure 19), the RNA velocity analysis also revealed genes previously unknown to have
588 dynamic behaviours across these stages; these included genes up-regulated from miracidia
589 tegument to sporocyst tegument, e.g. *epidermal growth factor receptor* (Smp_035260), genes
590 down-regulated from miracidia tegument to sporocyst tegument, e.g. *galectin domain*
591 (Smp_140590), and dynamic behaviour across stem cells and tegument, e.g. *unc44/ankyrin*
592 *2,3* (Smp_145700)(Supplementary Figure 19, Supplementary Table 16).
593

594 In the Kappa stem population, there was transcription and expression of genes detected in
595 germline stem cells (GSCs) and germ cells in the intra-mammalian stages of the life cycle
596 (Wang *et al.*, 2018; Li *et al*., 2021). A second P53 homolog, *p53-2* (Smp_136160), was
597 highly expressed and actively transcribed in Kappa cells (Figure 6F). In adults, this gene is
598 enriched in reproductive organs and has a genotoxic stress response that is thought to defend
599 the integrity of the genome by eliminating germ cells that have acquired mutations (Wendt *et*
600 *al*., 2022). Also expressed solely in Kappa cells (though some at low levels) are the germline-
601 specific regulatory genes *eled* (Smp_041540)(Figure 6G), *onecut1* (Smp_196950), *irx*
602 (Smp_063520), *akkn* (Smp_131630), and *boule* (Smp_144860)(expressed in the female
603 Kappa cells and Stem E, Supplementary Figure 22)(Wang *et al.*, 2018; Li *et al*., 2021), as

604 well as *klf* (Smp_172480) (Supplementary Figure 20), a homolog of KLF4, which is an
605 important transcriptional regulator governing pluripotency (Takahashi & Yamanaka, 2006).
606
607 Together, these data suggest that the Delta/Phi population is the origin of the tegumental
608 lineage and likely contributes to extra-embryonic somatic tissues during intra-molluscan
609 development, while the Kappa population likely contains pluripotent cells that will contribute
610 to the development of cercariae and the germline in intra-mammalian stages (Figure 7).
611
612 **Discussion**
613
614 Schistosomes are complex metazoan parasites, with adult worms of *Schistosoma mansoni*
615 characterised into 68 transcriptionally distinct cell clusters (Wendt *et al*., 2020).
616 Understanding the genes involved in the development and maintenance of these cell types is
617 a promising approach for identifying novel therapeutic targets (Wendt *et al*., 2020). Here we
618 have shown that the miracidium larva is composed of ~365 cells. Compared to later life cycle
619 stages, where the cercariae is made up of ~1,600 cells, the schistosomulum ~900 (Diaz *et al*.,
620 2020) and the adults tens of thousands of cells (unpublished data), the miracidium is a simple
621 stage of the life cycle. We identified and spatially resolved 19 transcriptionally distinct cell
622 types, revealing the relative contribution of each tissue type to the larva (Figure 8). The
623 miracidium, therefore, with its diversity of differentiated cells present in relatively low
624 numbers, makes a simple and tractable system with which to gain a fundamental
625 understanding of the biology and spatial architecture of schistosome cells, and their
626 transcriptomes.
627
628 Nineteen cell types may be a conservative characterisation, but at this resolution most cell
629 types formed clearly defined clusters. There was sufficient sensitivity to detect cell types with
630 as few as six cells per larva, for example, Neurons 2 and 3, and the protonephridial system.
631 Some rarer cell types are embedded in existing clusters, particularly in Neuron 1, and a
632 complete neural cell atlas and a study of the putative photoreceptor cells is in progress
633 (Rawlinson *et al*., in prep). Some cell types were preferentially recovered, e.g. the stem cells;
634 ISH revealed that they comprise only 7% of the total cell number in the larva, yet they
635 accounted for ~52% of the scRNA-seq data (Supplementary Table 17). Preferential recovery
636 may be due to differences in the robustness of different cell types to dissociation, cell sorting,
637 and/or the microfluidics of the 10X GEM capture process. The ciliary plates, for instance,

638 were only found in sufficient numbers to be defined as a cluster in unsorted samples.
639 Therefore, there are clear benefits to including unsorted cell samples to discover the diversity
640 of cell types in a tissue, organ, or organism, especially when cell types may be fragile, sticky,
641 or unusually shaped. Even with the inclusion of unsorted cells, some known cell types were
642 not captured, i.e. the penetration glands. With three gland cells per larva and a theoretical 56x
643 coverage of each cell, we would have expected ~168 gland cells in the dataset. These cells
644 were easily identified after dissociation by their distinct flask-shape but were not identifiable
645 in these 10X data. Perhaps, because of their large size (20x50µm) and irregular shape, they
646 were not incorporated into the 10X GEMs. Therefore, we used plate-based scRNA-seq for
647 the penetration glands to complete the cell type transcriptome atlas for all known cell types.
648 Together, we have transcriptomically characterised the anatomical systems described by Pan
649 (1980), and we discuss key findings below.
650
651 The two muscle clusters in our dataset correspond to the circular and longitudinal muscles
652 that comprise the orthogonal body wall muscles of the miracidium. We identified differences
653 between circular and longitudinal muscle transcriptomes that have also been observed in
654 planaria; e.g. *MyoD* and *Wnt-11-1* are expressed in longitudinal and circular muscles,
655 respectively (Scimone *et al*., 2017). Posterior Wnt signalling and anterior Wnt inhibition are
656 commonly observed features of anterior-posterior body patterning in animals (Peterson &
657 Reddien, 2009). The expression of *Wnt-11-1* and a Wnt inhibitor, *Notum,* in the circular
658 muscles at opposite poles of the miracidium suggested these muscles are involved in defining
659 the anterior-posterior axis. Furthermore, the unilateral, asymmetrical expression of *Notum* at
660 the anterior end is suggestive of a dorso-ventral axis and/or a left-right axis, neither of which
661 have been identified in the miracidium before. Many axial patterning genes expressed in
662 planaria muscles are essential for patterning during regeneration (Witchley *et al*., 2013; Vogg
663 *et al*., 2014). The expression of orthologs in *S. mansoni* muscles, a nonregenerative animal,
664 suggests that these genes may regulate stem cells during homeostasis in the adult (Wendt *et*
665 *al*., 2020). As the body axes of the miracidium were defined during embryogenesis, the
666 expression of these genes in the muscles of the larva suggest they will provide positional
667 information to the stem cells of the mother sporocyst in the axial patterning of the daughter
668 sporocysts.
669
670 The nervous system accounts for over 50% of cells that make up the larva and is the most
671 transcriptionally heterogeneous tissue, as observed in other animals (Siebert *et al*., 2019;

672 Vergara *et al*., 2021). At the end of embryogenesis, at hatching, most of the miracidia
673 neurons are fully differentiated and form the functional neural circuits of the larva. This is
674 evident in our scRNA-seq data by the presence of discrete, well-defined neural cell clusters.
675 *In situ* validation of marker genes revealed many of the nerves identified ultrastructurally; for
676 example, the location and shape of Neuron 4 cells resemble that of the bipolar ganglion cells
677 with multi-ciliated nerve endings described by Pan, 1980. Examination of their
678 transcriptional profiles indicates their function; top markers include a putative *mTORC1*, two
679 *Polycystin* genes, a *GTP-binding ras-like* and *fibrocystin.* Mammalian orthologs of these
680 function in cells with primary cilia in the detection of flow sheer and regulate mTORC1
681 signalling to maintain the cell in a quiescent and functional state (Lai & Jiang, 2020).
682 Polycystins 1 and 2 expressed in ciliated neurons of the larva of the annelid *Platynereis*
683 *dumerilli* are involved in mechanosensation, detection of water vibrations and mediate the
684 startle response (Bezares-Calderón *et al.,* 2018). Furthermore, in planaria, SoxB1-2 regulates
685 polycystin genes and together they function in ciliated cells that are involved in water flow
686 sensation (Ross *et al*., 2018). The gene expression profile of Neuron 4 cells in the miracidium
687 suggests they could function as flow and/or vibration sensors and may be specified by a Sox
688 transcription factor. These multiciliated receptors degenerate 12 hours after the miracidia
689 transform into a mother sporocyst (Poteaux *et al*., 2023).
690
691 *In situ* validation of marker genes also revealed an undescribed cell type, namely the six
692 submuscular cells that express the top marker genes for the ciliary plates (i.e. a 16KDa
693 calcium binding protein and *p25 alpha*). The expression of these genes in both the ciliary
694 plates and the submuscular cells suggests that these latter cells are transcriptionally similar
695 and might play a role in the development and functioning of the ciliary plates. Their
696 projection into the external environment suggests a sensory role, which, together with their
697 location between the first and second tier of ciliary plates, could suggest a role in
698 coordinating the movement of the motile cilia across the plates and tiers in response to
699 sensory cues (i.e. similar in function to the ciliomotor pacemaker neurons in *Platynereis*
700 larvae, Verasztó *et al*., 2017). The ciliary plates are a miracidium-specific cell type (it is the
701 only life cycle stage that uses motile cilia for locomotion), and once inside the snail, these
702 plates are shed within 2 – 48 hours (Meulement *et al*., 1978; Poteaux *et al*., 2023). As
703 expected, they are transcriptomically typified by enrichment with dynein and high metabolic
704 activity. The 16KDa calcium binding protein (Smp_096390) is clearly important to the
705 function of ciliary plates as it is highly expressed in data collected from both droplet- and

706 plate-based scRNA-seq platforms. Accordingly, bulk RNAseq data shows this gene was
707 previously detected in the late embryonic stage, miracidia and early mother sporocyst stage
708 (Lu *et al*., 2021; Buddenborg *et al*., 2023).
709
710 Between all the ciliary plates sit the ridges of the tegumental cells. The ridges connect to the
711 cell bodies beneath the body wall musculature, and their transcriptomes show a predicted
712 network of genes involved in protein folding, export and translocation. This is consistent with
713 ultrastructural studies that have observed abundant RER, ribosomes, Golgi apparatus, and
714 complex carbohydrates in these cell bodies, suggesting they are active in producing
715 exportable proteins (Meulement *et al*., 1978; Pan, 1980). Large numbers of membrane-bound
716 vesicles were also observed and thought to become part of the new tegument membrane
717 (syncytium) that rapidly forms around the mother sporocyst during the early stages of
718 transformation, protecting the mother sporocyst from attack by the snails' amoebocytes (Pan,
719 1980). The RNA velocity analysis has revealed dynamic genes that may be crucial in the
720 maturation of the tegument cells across this life cycle transition, and in turn, for evasion of
721 host defences and parasite survival. This analysis also suggests a developmental trajectory for
722 the tegument cells from the miracidia Delta/Phi stem cells and follows their progeny to
723 terminally differentiated tegument cells in the miracidium and five-day-old sporocyst. Many
724 of the molecular regulators and markers of adult tegument biogenesis (Collins *et al*., 2016;
725 Wendt *et al*., 2018; Wendt *et al*., 2022) are expressed in the miracidia and sporocyst stem and
726 tegument cells, indicating that there is a conserved molecular program for the development of
727 the syncytial tegument in the mother sporocyst and its maintenance in the adult worm.
728
729 The transcriptomes of the two stem populations in the miracidium suggest they are in
730 different activation states. This finding aligns with the discovery of two stem cell populations
731 in the mother sporocyst that have different proliferation kinetics (Wang *et al*., 2013). The pre-
732 activated state of Delta/Phi stem cells in the miracidia may signify that they are the first
733 population to proliferate once the miracidium transforms into the mother sporocyst inside the
734 snail. These two stem populations may also have different potency states; the co-expression
735 of multiple transcription factors in Delta/Phi, including those specific to the tegument lineage
736 (Wendt *et al*., 2018; 2022), indicates a multipotent progenitor stage. Whereas in the Kappa
737 stem cells, the expression of genes that define the germline in the intra-mammalian stages
738 (Wang *et al*., 2018) suggest pluripotency. Supporting this, the RNA velocity analysis shows
739 that *p53-2*, the genome protection gene (Wendt *et al*., 2022), is most highly expressed, and

740 being actively transcribed, in the Kappa population. The incorporation of miracidia single-
741 cell data into existing stem cell scenarios on *Schistosoma mansoni* development (Wang *et al*.,
742 2013; Wang *et al*., 2018; Li *et al*., 2021) shows the continuum of the Kappa population from
743 hatching through the intra-molluscan stages to the adult stage inside the mammalian host
744 (Figure 7). The finding that *eled*, and many of the germline-specific regulatory program
745 components, are expressed in some miracidia Kappa cells suggests that when a miracidium
746 hatches, the cells that will contribute to the germline may already be segregated and the
747 molecular regulatory program that differentiates somatic (Delta/Phi) and germ cell (Kappa)
748 lineages is present. Intriguingly, the presence of spliced-only copies of the germline defining
749 genes *eled* and *boule* could suggest that they are maternal transcripts that have been restricted
750 to the primordial germ cells during embryogenesis, as is the case in Zebrafish embryos
751 (Fishman et al., 2023). An alternative explanation is that unspliced transcripts exist for these
752 lowly expressed genes but their abundance was below our threshold for detection.
753
754 An unexpected finding regarding the stem cells was the further clustering of the two
755 populations by sex. This is the first identification of any sex differences in the miracidium.
756 This degree of transcriptomic sexual difference was most evident in the Delta/Phi and Kappa
757 stem cell clusters, with no distinct clustering by sex observed in other cell clusters. This could
758 be explained by incomplete gene dosage compensation in the miracidium. Following
759 investigations of differences in gene dosage compensation across developmental stages using
760 bulk RNA-seq (Picard *et al*., 2019), our study suggests tissue-specific variation could add a
761 further dimension to the role of dosage compensation in schistosomes. A similar phenomenon
762 has been described in mouse embryonic stem cells and in *Drosophila,* where a precise
763 regulation of a gene-by-gene dosage compensation mechanism is critical to the proper
764 development of specific tissues and organs, such as the wing and eye (Valsecchi *et al*., 2018).
765 In contrast to the miracidium, sex differences in undifferentiated stem cells have not been
766 reported in later stages of the schistosome life cycle (Wang *et al.,* 2018; Diaz *et al.*, 2020;
767 Wendt *et al*., 2020). In fact, it may not have been possible to detect these differences in
768 earlier studies because scRNA-seq data were mapped onto older versions of the genome,
769 where the sex chromosomes were not as well defined (Buddenborg *et al*., 2021). In addition
770 to using an improved reference annotation, the present study has also benefited from a high
771 level of coverage, enabling differences in stem cells to be more easily discerned. Whether
772 these sex-linked gene expression and dosage differences have functional implications in
773 developing male and female sporocysts remains to be investigated.

774 The current understanding of schistosome development is that differentiated somatic cells of
775 the miracidium do not continue into the daughter sporocyst stage, the cercariae and beyond;
776 it is the stem cells that persist and differentiate into the cell types of the subsequent life stages
777 (Wang *et al*., 2018). In this respect, all miracidium somatic cells are miracidium-only cells.
778 Our atlas shows that their transcriptional signatures are similar to comparable tissue types at
779 later stages (Diaz *et al*., 2020; Wendt *et al*., 2020, Diaz & Attenborough *et al*., 2024) and that
780 there may be conserved tissue-specific molecular developmental programs that are deployed
781 multiple times during the complex life cycle. Through this atlas, we have shown that even in
782 this simple and sexually monomorphic short-lived stage, there are different populations of
783 muscle, a diverse and distinct range of neural cell types, and that stem cell transcriptomes are
784 already different based on the sex of the miracidium. The molecular definitions and spatial
785 analyses of the miracidia cell types we have presented here allow us to explore the molecular
786 processes and pathways to reveal cell functions and interactions. In addition, high coverage
787 data from individual cells has allowed us to examine how gene expression changes across
788 and within tissue types. Already, a picture is emerging of how the peripheral neurons may
789 detect environmental cues, connect to the brain, muscles and ciliary plates to orientate in the
790 water column; how the parenchymal cells supply nutrients to the surrounding cells; what the
791 contents of the penetration glands may be; and how the tegumental and stem cells are primed
792 for the transition to the intra-molluscan stage.

793

794 **Methods**

795 **Parasite material**

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

803

804 **Nuclei count**

805 *Schistosoma mansoni* eggs isolated from infected mouse livers were hatched under light in
806 freshwater (Mann *et al.,* 2010), and the miracidia collected within four hours of hatching. The
807 sample was put on ice for an hour to sink and concentrate the larvae, and all but 1 mL of the
808 water was removed. One mL of 8% paraformaldehyde (4% final PFA concentration) was
809 added to fix the miracidia for 30 minutes at room temperature. The larvae were rinsed in
810 PBS/Tween (0.1%) and mounted in DAPI Fluoromount for confocal imaging. The whole
811 mounts were imaged on a Leica Sp8 using a 40X objective with the pinhole set to 1AU, and a
812 30-50% overlap between Z slices. The Z-stacks were imported into ImageJ, and each nucleus
813 was manually segmented and counted using the plugin TrakEM2 (Cardona *et al*., 2012).
814
815 **Single-cell tissue dissociation**
816 *Schistosoma mansoni* miracidia for single-cell dissociation were collected as described
817 above. Eggs isolated from mouse livers were hatched in nuclease-free water (NFW) and
818 collected within four hours of hatching. Two samples of miracidia were collected; sample 1
819 contained approximately ~8250 individuals, and sample 2 ~11,700. The samples (live
820 miracidia in NFW in Low Protein Binding 15 mL Falcon Tubes [cat# 30122216]) were
821 placed on ice for an hour to sink and concentrate the larvae, and all but 20 µl of water was
822 then removed.
823
824 Three hundred µl of digestion buffer (2.5 µl of 0.5 M EDTA in 300 µl of Hank's Buffered
825 Salt Solution without Ca2+ and Mg2+ [HBSS-/-]) was added to the larvae. We reconstituted
826 250 mg of the enzyme Native *Bacillus licheniformis* Protease (NATE-0633, Creative
827 Enzymes)(Adam *et al*., 2017) in 2.5 mL of HBSS-/- (to give a stock concentration of 100
828 mg/ml) and added 200 µl of this stock to the larvae in buffer (final enzyme concentration =
829 40 mg/ml). The samples were kept on ice for 45 minutes and pipetted gently ~20 times every
830 five minutes. Two hundred µl of 25% BSA in 1X HBSS-/- was added to the sample to
831 quench the enzyme, and 10 ml of ice-cold PBS +/+ was added to bring the BSA
832 to ~0.5%. The samples were centrifuged at 350 g for 5 minutes at 4°C, and all but 20 µl of
833 the supernatant was removed. We reconstituted 5 mg of the enzyme Liberase™ (Thermolysin
834 Medium) Research Grade (LIBTM-RO Roche) in 2 mL of PBS (to give a stock concentration
835 of 2.5 mg/ml) and added 100 µl of this stock to the sample (final enzyme concentration =
836 2.08 mg/ml). The samples were incubated at room temperature for 15 minutes,
837 pipetting gently ~10 times every five minutes. The suspension was passed through 60 µm and
838 30 µm cell strainers (Falcon) and resuspended in 5 mL of cold 1× PBS supplemented with

839 0.5% BSA. The samples were centrifuged at 350 g for 10 minutes at 4°C, and all but 700 µl
840 of the supernatant was removed.
841
842 Sample 1 cell suspension was sorted; 200 µl was used for a no-stain control for
843 autofluorescence during cell counting and sorting. The remaining 500 µl of cell suspension
844 was stained with 1 µg/ml of Propidium Iodide (PI; Sigma P4864) to label dead/dying cells.
845 Sample 1 was sorted into an eppendorf tube using the BD Influx™ cell sorter (Becton
846 Dickinson, NJ) by enriching for PI negative cells (i.e., live cells). It took ~3 hours from the
847 beginning of the enzymatic digestion to generating single-cell suspensions ready for library
848 preparation on the 10X Genomics Chromium platform. Sample 2 cell suspension was not
849 sorted.
850
851 **10X Genomics library preparation and sequencing**
852 The 10X Genomics protocol ("Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User
853 Guide" available from
854 https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f
855 9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf) was followed
856 to create gel in emulsion beads (GEMs) containing single cells, hydrogel beads and
857 reagents for reverse transcription, perform barcoded cDNA synthesis, and produce
858 sequencing libraries from pooled cDNAs. The concentration of single-cell suspensions
859 was approximately 119 cells/µl (sample 1) and 122 cells/µl (sample 2), as estimated by flow
860 cytometry-based counting (Beckman Coulter Cytoflex S). We concentrated both samples to
861 732 cells/µl. We loaded two aliquots of sample 1 and two of sample 2 into four 10X reactions
862 according to the 10X protocol (Chromium Single Cell 3' Reagent Kits v3.1), intending to
863 capture approximately 10,000 cells per reaction.
864
865 Library construction (following GEM breakage) using 10X reagents following the "Single
866 Cell 3'Reagent Kits v3.1 User Guide". The four libraries were sequenced on one lane of
867 Illumina Novaseq XP (28 bp read 1, 90 bp read 2, 10 bp index 1, 10 bp index 2). All raw
868 sequence data will be deposited in the ENA.
869
870 **Mapping and quantification of single-cell RNA-seq**
871 We mapped the single-cell RNA-seq data to version 10 of the *S. mansoni* reference genome
872 (Buddenborg *et al*., (2021) and WormBase ParaSite ((Howe *et al*., 2016; 2017)) using 10X

873 Genomics Cell Ranger (version 6.0.1) (Zheng *et al*., 2017). Before mapping, exon eight was
874 removed from Smp_323380 and the 3' end of was trimmed from Smp_337410 to reduce
875 erroneous mapping to the W chromosome from Z gametologs. The Cell Ranger
876 implementation of EmptyDrops (Lun *et al*., 2019) was used to detect cells and empty
877 droplets. Across the four samples, on average 63.88% of the reads mapped confidently to the
878 transcriptome (63.4% for sorted, 64.35% for unsorted), and an average 83.65% of reads were
879 associated with a cell. There were an average of 118,395 reads per cell. Across the four
880 samples, we captured an estimated 33,319 cells: 8,815 and 24,504 cells from sorted and
881 unsorted samples, respectively. The median gene count was 1,660, and the median UMI
882 count was 5,343. Following QC, 20,478 cells remained for onward analysis.
883
884 **Quality control (QC) and clustering using Seurat**
885 The four samples were imported into Seurat (version 4.3.0) (Hao *et al*., 2021) using R
886 (version 4.1.3) in RStudio (version 1.4.1106) for quality control, clustering, and marker gene
887 identification. The four samples were initially processed separately. Cells with less than 200
888 genes expressed and genes in less than three cells were removed. QC metrics, including gene
889 distribution and UMI counts per cell, were inspected visually and used to select QC cut-offs.
890 Following a preliminary analysis without further cell filtering, cells with fewer than 500
891 UMIs were removed, and a maximum mitochondrial threshold of 5% was applied. This was
892 based on data distribution, best practices from the authors of the software, and other similar
893 studies; here, we were aiming to remove low-quality cells while being relatively permissive
894 as there may be cell types in these data outside of standard filtering parameters. Based on
895 existing knowledge of this larva, we were expecting both multinucleate and anucleate cell
896 type(s) as well as single-nuclei cells (Pan, 1980). The data were normalised, scaled
897 (regressing for mitochondrial gene percentage), and variable genes were identified for
898 clustering following Seurat guidelines. Throughout the analysis of these data, several tools
899 were used to select the optimal number of principal components (PCs) to use: Seurat's
900 ElbowPlot and JackStraw, and molecular cross-validation
901 (https://github.com/constantAmateur/MCVR/blob/master/code.R), and Seurat's clustree tool
902 was used to assess the stability of clusters with increasing resolution. We employed
903 DoubletFinder (McGinnis *et al*., 2019) to identify multiplets, which can lead to spurious
904 signals. The estimated doublet rate was calculated based on 10X loading guidelines and
905 adjusted for our sample concentrations; we estimated multiplet rates at 8% for sorted samples
906 and 52.8% for unsorted samples. We used paramSweep, as recommended, to identify the

907 optimum parameters for each sample in our data and removed cells classified as multiplets.
908 DoubletFinder removed 299 and 303 cells from the two sorted samples, and 6,069 and 5,496
909 cells from the two unsorted samples.
910
911 We then combined the two sorted samples and the two unsorted samples (producing one
912 sorted object and one unsorted object). We used SCTransform to normalise and scale these
913 two objects (with regression for mitochondrial percentage and sample ID). Seurat's
914 IntegrateData function was then used to integrate the sorted and unsorted samples (with
915 normalization.method = "SCT"). This resulted in 20,478 cells passing all the QC filters and
916 going onto the primary analysis, 3,849 and 3,962 from the sorted samples and 6,702 and
917 5,965 from the unsorted samples.
918
919 Following integration, principal component analysis (PCA) was run for up to 100 PCs. The
920 data were explored using a range of PCs and resolutions and using the tools listed above, 55
921 PCs and a resolution of 1 were selected for the analysis. We manually segmented out the cells
922 in the Neuron 3 cluster based on clear differences in gene expression. Top markers were
923 identified using Seurat's FindAllMarkers (min.pct = 0.0, logfc.threshold = 0.0, test.use =
924 "roc") as recommended by Seurat best practices (https://satijalab.org/seurat/) to capture subtle
925 differences in gene expression, and were sorted by AUC. We used the identity and expression
926 of top marker genes (as identified by AUC), along with previously published cell and tissue
927 type marker gene expression, to determine the identity of the cell clusters. To link gene
928 description and chromosome locations to gene IDs, the GFF for the latest publicly available
929 genome version was downloaded from WBPS (WBPS18 release, Howe *et al*., 2016; 2017)).
930
931 **Sex-effect analysis**
932 To investigate sex effects on the single cell transcriptome, the cells were categorised as W+
933 or W- (an approximation of female and male cells respectively, as males lack WSR genes).
934 As ground-truth data was absent, the % of reads mapping to ZSR and WSR was calculated
935 using Seurat's PercentageFeatureSet. Cells with 0.2% or more reads mapping to the WSR
936 were classified as W+, and the remaining cells were classified as W-. The 0.2% threshold
937 was selected (Supplementary Figure 14) to avoid classifying ZZ cells with a few transcripts
938 erroneously mapping to a gene on the WSR as W+. This produced 8,702 cells labelled as
939 W+, and 11,776 cells labelled as W-.
940

941 **Self-Assembling Manifold (SAM) algorithm**

942 The count data from the QC-filtered Seurat analysis was exported for SAM (v1.0.1) analysis.
943 For each SAM analysis, the raw count data was imported as a SAM object, standard
944 preprocessing done (using default SAM parameters), and the SAM algorithm (Tarashansky *et*
945 *al*., 2019) run. For further analysis, the object was converted into a Scanpy (v 1.8.2) (Wolf *et*
946 *al*., (2018)) object, top markers extracted, and metadata from the Seurat analysis was
947 attached. This method was repeated on the Neuron 1 cluster alone, to better understand the
948 transcriptome heterogeneity.
949
950 **RNA velocity**
951 All four miracidia samples and the raw data from the mother sporocyst (Diaz &
952 Attenborough *et al*., 2024) were pseudo-mapped using kallisto bustools (v 0.27.3) (Melsted *et*
953 *al*., 2021) using the lamanno workflow to the version 10 *S. mansoni* genome (WBPS18
954 release (Howe *et al*., 2016; 2017, with the same amendments as listed above). The filtered
955 spliced and unspliced matrices for each sample individually were imported to R. DropletUtils
956 (v 1.20.0) (Lun *et al*., 2019) and BUSpaRse (v 1.14.1) (Moses and Pachter, 2023) were used
957 to process and QC the matrices, and filter out cells and genes with few reads. The matrices
958 were then converted to Seurat objects and further filtered so only cells with the cell barcodes
959 that passed QC filtering in the primary Seurat analysis were retained. The metadata and raw
960 counts of the unspliced and spliced reads following filtering were exported as CSV files. The
961 same method was performed on the mother sporocyst samples, which were combined before
962 exporting. For the joint analyses of miracidia and sporocyst cells, the combined sporocyst
963 Seurat object was merged with each of the four miracidia samples separately (due to the
964 mismatch in sample size and batch effect between sorted and unsorted samples) and all were
965 exported.
966
967 The CSV files containing spliced and unspliced reads, together with sample metadata were
968 combined to make an AnnData object using Scanpy (v 1.8.2, Python v 3.8.5) (Wolf *et al*.,
969 2018). Using scvelo (v 0.2.4) (Bergen *et al*., 2020), the standard parameters were used to
970 filter and normalise the data, compute moments, and recover dynamics. Velocity was
971 calculated using the dynamical mode. Following PCA, a UMAP was generated
972 (n_neighbors=10, n_pcs=40) and Leiden clustering was added (resolution=0.8). The top-
973 ranked dynamical genes for each cluster and known genes of interest were plotted. This was

974 performed for each of the four miracidia samples combined with the sporocyst samples, with
975 and without the use of combat for batch correction. As the sporocyst samples were also
976 sorted, we used the sorted miracidia samples to reduce technical effects.
977
978 To focus on the stem/germinal and tegument clusters, the same method as above was used
979 after subsetting the data in Seurat to contain only cells from the clusters Stem/germinal,
980 Tegument 1, and Tegument 2 in the sporocyst samples, and only Stem A-G and Tegument
981 clusters from the miracidia clusters.
982
983 We present the stem and tegument cells from the sporocyst stage merged with sample 1
984 (sorted). The equivalent results from sporocysts merged with sample 4 of the miracidia and
985 results with all cells from those two conditions are available in the supplementary data
986 (Supplementary Figures 21-23).
987
988 **EdU-staining experiment**
989 Miracidia were collected as described above for 2 hours. As miracidia continually hatched
990 over this time period, they ranged from 0 to 2 hours post-hatching. The miracidia were
991 collected and split into two samples: one for EdU labelling and the second for a negative
992 control (miracidia in water without EdU). Sample one was pulsed with 500 µM EdU in water
993 for 4 hours. Sample two was kept in water for 4 hours. Both samples were fixed in 4% PFA
994 for 30 minutes at room temperature. EdU incorporation was detected by click reaction with
995 25 µM Alexa Fluor488 Azide conjugates (Invitrogen) for 30 minutes. As a positive control,
996 we pulsed freshly-transformed mother sporocysts in EdU for 3 days (Wang *et al*., 2013).
997 Three biological replicates were carried out, with 20 miracidia and 20 mother sporocysts per
998 treatment examined for each replicate.
999
1000 **Plate-based scRNAseq analysis of individually isolated cells**
1001 A sample of thousands of miracidia was dissociated into a single cell suspension using the
1002 dissociation protocol described above. Individual gland cells, ciliary plates and unknown
1003 cells were picked up one at time using an EZ Grip transfer pipette set to aspirate a 3µl
1004 volume (Research Instruments Ltd 7-72-2800). Each cell was placed in a well of a 96-well
1005 plate containing a lysis buffer (NEBNext Cell Lysis Buffer, E6428). cDNA was generated
1006 using the NEB low input RNA kit (E6420). DNA library construction was carried out using
1007 the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (E7805). The libraries were

1008 sequenced on one lane of NovaSeq SP. The reads from the 96 samples were pseudoaligned to
1009 the version 10 *S. mansoni* genome (WBPS18 release (Howe *et al*., 2016; 2017, with the same
1010 amendments as listed above) using kallisto (v 0.44.0) (Bray *et al*., 2016)). These were
1011 imported into R (v4.3.1) and the metadata were inspected, after which only samples coming
1012 from a single cell with a minimum of 20% reads pseudoaligned were retained (n=76). The
1013 samples were then processed using sleuth (v0.30.1) (Pimentel *et al*., 2017) following
1014 guidance from sleuth tutorials, including the steps for normalising, fitting error models, and
1015 differential analysis. Differential analysis was used to identify significant genes between the
1016 cell types, which were visualised, along with known genes of interest, using sleuth's heatmap
1017 function. Furthermore, the kallisto outputs were used to calculate the top 200 genes by
1018 median transcripts per million (TPM) of the two key cell types of interest, ciliary plates and
1019 gland cells, and these were used for GO term analysis following the method described below.
1020

1021 ***In situ* hybridisation**

1022 Miracidia collected as described above within 4 hours of hatching were fixed in 4% PFA in 2
1023 µm filtered PBSTw (1x PBS + 0.1% Tween) at room temperature for 30 minutes on a rocker.
1024 The larvae were transferred to incubation baskets (Intavis, 35 µm mesh) in a 24-well plate,
1025 rinsed 3 x 5 minutes in PBSTw, and incubated in 5 µg/ml Proteinase K (Invitrogen) in 1x
1026 PBSTw for 5 minutes at room temperature. They were post-fixed in 4% Formaldehyde in
1027 PBSTx for 10 minutes at room temperature and then rinsed in PBSTw for 10 minutes. From
1028 this point on, the *in situ* hybridisation experiments followed the protocol described by Choi *et*
1029 *al*., (2016; 2018) and developed for wholemount nematode larvae. We carried out fluorescent
1030 *in situ* hybridisation on 27 cell-type marker genes. Probes, buffers, and hairpins for third-
1031 generation *in situ* hybridization chain reaction (HCR) experiments were purchased from
1032 Molecular Instruments (Los Angeles, California, USA). Molecular Instruments designed
1033 probes against the sequences on WormBase ParaSite (WBPS version 18, WS271, Howe *et*
1034 *al*., 2017)(Supplementary Table 18). We subsequently labelled some larvae with phalloidin
1035 by adding 1 µl phalloidin to 1 ml 5X SSCT for an hour at room temperature in the dark, then
1036 rinsing 6 x 30 minutes in SSCT, incubating overnight at 4°C in DAPI fluoromount-G
1037 (Southern Biotech) and then mounting and imaging on a confocal laser microscope (Leica
1038 Sp8). We carried out at more than three *in situ* hybridisation experiments for each marker
1039 gene we validated (each experiment was a biological replicate). From each experiment we
1040 imaged (by confocal microscopy) 10 miracidia (technical replicates) per marker gene
1041 experiment. To count the number of cells belonging to each tissue type we imported confocal

1042 Z-stacks of tissue marker gene expression (Supplementary Videos 18-20) into ImageJ, and
1043 using the plugin TrakEM2 (Cardona *et al*., 2012), manually segmented each nuclei that was
1044 expressing, and was surrounded, by the transcripts.
1045

1046 **Opsin phylogeny**

1047

1048 Amino acid sequences for selected metazoan opsin proteins were obtained from the GenBank
1049 database and aligned using the —auto mode of MAFFT v7.450 (Katoh & Standley, 2013),
1050 then trimmed using trimAl v1.4.rev22 (Capella-Gutierrez *et al*., 2009) to remove alignment
1051 columns with more than 50% gap characters ensuring at least 50% of alignment columns
1052 were maintained in the final alignment. Phylogenetic analysis of this alignment was
1053 performed using RaxML v8.2.12 (Stamatakis, 2014). The best-fitting empirical amino acid
1054 substitution model was identified as maximising the log-likelihood of the data under a model
1055 incorporating gamma-distributed variation in substitution rates across sites, and this optimal
1056 model (LG) was used for subsequent inference. The maximum-likelihood tree and branch
1057 lengths were identified as the highest likelihood from 25 independent searches starting from
1058 different random starting tree topologies, and support for partitions on this tree was estimated
1059 using 1000 bootstrap replicates of resampling alignment positions with replacement, using
1060 the fast bootstrap (-x) algorithm implemented in RaxML.
1061

1062 **Protein-Protein interaction analysis**

1063 Protein interactions were predicted using the online search tool STRING (www.string-db.org;
1064 V 11) (Szklarczyk *et al.*, 2019) for the following well-defined cell clusters: Neuron 3-5,
1065 ciliary plates, protonephridia, Parenchyma 1 & 2, tegument and the two main stem cell
1066 populations, Delta/Phi and Kappa. Protein sequences for the top 100 marker genes for each
1067 cell cluster were collected from WormBase ParaSite. The protein sequences were entered as a
1068 multiple protein search. Default settings were used to predict interactions with either a
1069 minimum interaction (confidence) score of 0.4 or 0.7, corresponding to a medium or high
1070 level of confidence.
1071

1072 **Gene Ontology enrichment analysis**

1073 Each cluster's list of marker genes was filtered to focus on specific cluster markers. Genes
1074 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. The same method was used to identify overrepresented terms in the WSR genes, ZSR genes, and WSR together with ZSR genes for the BP, MF and CC categories.

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. All supplementary material (figures, tables and videos) is available on BioRxiv.

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 analysed data
KR performed ISH experiments
SD annotated marker genes in v10 of the genome

1109 TA and KR drafted the complete manuscript, with sections of text contributed by GR and
1110 MB
1111 JC carried out the opsin phylogeny
1112 All authors read and approved the final manuscript.
1113
1114 **Acknowledgements**
1115 We thank Simon Clare and his team for technical assistance with the life cycle maintenance
1116 and collection of parasite material at the Wellcome Sanger Institute, and Karl Hoffmann,
1117 Josephine Fforde-Thomas and Benjamin Hulme for additional ongoing parasite material. We
1118 thank Sarah Buddenborg for work on the annotation of the *Schistosoma mansoni* genome;
1119 David Goulding for assistance with Microscopy; teams within DNA Pipelines for data
1120 production; and core Informatics teams and Pathogen Informatics for the infrastructure used
1121 for data analysis. We thank members of the Parasite Genomics team at the Wellcome Sanger
1122 Institute for their comments and input on this study. Finally, we thank the reviewers and the
1123 editor for their valuable feedback on this manuscript.
1124
1125
1126 **Figure legends**
1127
1128 **Figure 1**. **Identification of 19 transcriptionally distinct cell types in the miracidium.**
1129 **A)** The miracidium is composed of ~365 cells; **i)** DIC image of miracidium, **ii)** a Maximum
1130 Intensity Projection (MIP) of a confocal z-stack of DAPI-stained miracidium with nuclei
1131 segmented to enable counting (larval anterior pole at the top in all images). **B)** Experimental
1132 scheme describing the parasite's life cycle (images of developmental stages not to scale),
1133 parasite dissociation, single-cell analysis of miracidium and validation pipeline. An average
1134 of 9975 miracidia per sample were dissociated; two samples were enriched for live cells
1135 (propidium iodide negative) using fluorescence-activated cell sorting (FACS), another two
1136 samples were unsorted. Cells were loaded according to the 10X Chromium single-cell 3'
1137 protocol. Clustering was carried out to identify distinct populations and population-specific
1138 markers. Validation of population-specific markers was performed by *in situ* hybridisation
1139 (ISH). **C)** Uniform Manifold Approximation and Projection (UMAP) representation of
1140 20,478 miracidium single cells. Cell clusters are coloured and distinctively labelled by cluster
1141 identity. **D)** Gene-expression profiles of the top population markers identified for each cell
1142 cluster (gene identifiers shown in parenthesis but with "Smp_" prefix removed for brevity).

1143 The colours represent the expression level from yellow (high expression) to dark blue (low
1144 expression). Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0). The
1145 sizes of the circles represent the percentages of cells in those clusters that expressed a specific
1146 gene.
1147
1148 **Figure 2. Orthogonal body wall muscles are transcriptionally distinct. A)** Dot plot
1149 depicting the expression profiles of specific- or enriched- marker genes in the two muscle
1150 clusters. Genes validated by *in situ* hybridisation (ISH) are marked in red. Gene identifiers
1151 shown in parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been
1152 log-normalised and scaled using Seurat(v. 4.3.0). **B) i-ii)** Wholemount ISH (WISH) of
1153 paramyosin *PRM*$^+$ and counterstaining with phalloidin reveals the nuclei of the circular body
1154 wall muscles (BWM), which form two distinct bilaterally symmetrical lines that run
1155 peripherally from the pole of the larva to the other (arrowheads in ii). The longitudinal BWM
1156 nuclei (ii - arrows) are spaced regularly between the lines of the circular BWM nuclei and
1157 their actin fibres run orthogonally to the circular muscles. These patterns were seen in 100%
1158 of individuals examined, n = 50. **C) i)** Segmentation of the *PRM*$^+$ cells in one miracidium
1159 shows 74 muscle cells in total: ~28 circular (segmented in orange), 33 longitudinal (yellow)
1160 and **ii)** 13 in an anterior unilateral cluster (inside dashed cyan line) adjacent to the apical
1161 gland cell (identifiable by its four nuclei, inside dashed red circle). **D)** Expression of markers
1162 for muscle 1 cluster; **i)** *Kunitz-type protease inhibitor* is expressed in circular BWM and *wnt-*
1163 *11-1* is expressed in 7 *Kunitz-type protease inhibitor*$^+$ cells at the posterior pole; **ii)** close-up
1164 of posterior end of another miracidium showing the co-expression of *Kunitz-type protease*
1165 *inhibitor* and *wnt-11-1.* In 100% of individuals examined, n = 30. **E) i-iii)** *Notum,* an inhibitor
1166 of *wnt* signalling, is expressed highly at the opposite pole to *wnt-11-1* in the cluster of 13
1167 muscle nuclei adjacent to the apical gland cell (nuclei in dashed red circle). In 100% of
1168 individuals examined, n = 30. **F)** Expression of markers for muscle 2 cluster; a *calcium*
1169 *activated potassium channel* (Smp_156150) and the transcription factor *myoD* (Smp_167400)
1170 show co-expression in longitudinal BWMs (arrowheads). In 100% of individuals examined, n
1171 = 30. Scale shown in B also applies to C, Di, E and Fi.
1172
1173
1174 **Figure 3. Neural complexity and ciliary plates in a simple larva. A)** Expression profiles of
1175 cell marker genes that are specific or enriched in the five neuronal clusters and ciliary plates.
1176 Genes validated by *in situ* hybridisation (ISH) are marked in red. Gene identifiers shown in

1177 parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been log-
1178 normalised and scaled using Seurat(v. 4.3.0). *Complexin* (*cpx*) is expressed in all five nerve
1179 clusters and is a neural marker in the adult (Wendt *et al*., 2020). **B) i)** A dorso-ventral view
1180 of WISH of *cpx* shows expression in the brain, to the anterior - with projections around the
1181 three gland cells, and in two posterior clusters (100% of individuals examined, n = 50). **ii)**
1182 Segmentation of *cpx*$^{+}$ cells (in magenta) in one miracidium (lateral view) reveals that at least
1183 209 of the 365 cells that make up the larva are neural and that the clusters of *cpx* cells that are
1184 located posterior to the brain, and send projections to the posterior pole, sit perpendicular to
1185 the circular muscle cell nuclei (blue dashed lines). **C)** Multiplexed ISH for top markers for
1186 neuron 2 (Smp_201600) and neuron 3 (Smp_071050) shows expression in multiple cells of
1187 the brain, including cells adjacent to each other. Smp_071050 is also expressed in the apical
1188 gland cell (red dashed circle). 100% of individuals examined, n = 30. **D)** The top marker for
1189 neuron 4 Smp_319480 is expressed in four cells anteriorly (arrows) and ~22 that sit lateral
1190 and just posterior to the brain; these later cells send projections into the brain and out to the
1191 body wall at the latitude between the 2nd and 3rd tier of ciliary plates **(iii)** seen here expressing
1192 the ciliary plate marker gene Smp_096390. 100% of individuals examined, n > 30. **E)** Neuron
1193 5 marker Smp_343710*,* an EF-hand domain-containing protein, is expressed in **(i)** 10-20 cells
1194 whose nuclei sit outside of, and anterior to, the brain, in and around the paired lateral glands
1195 and their secretory ducts, and **(Eii & iii, F)** in a pair of bilaterally symmetrical bulbous
1196 protrusions. 100% of individuals examined, n = 30. **F)** Summary schematic of the *in situ*
1197 expression of marker genes for the neural cell clusters. **G) i)** A top marker for ciliary plates, a
1198 calcium binding protein (Smp_096390), shows transcripts expressed in all the ciliary plates
1199 of all four tiers (brackets) and in 6 cells towards the anterior pole. **(ii & iii)** Counterstaining
1200 with phalloidin shows that the nuclei of these anterior cells sit beneath the body wall muscle
1201 and send a projection externally between the 1st and 2nd tier of ciliary plates (arrowheads).
1202 100% of individuals examined, n > 30. Scale shown in B also applies to C, D, Ei and Gi.
1203
1204
1205 **Figure 4. Identification of the tegument, protonephridia, parenchymal cells in the**
1206 **miracidium**. **A)** Expression profiles of cell marker genes specific or enriched in these cell
1207 type clusters. Genes validated by *in situ* hybridisation are in red. Gene identifiers shown in
1208 parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been log-
1209 normalised and scaled using Seurat(v. 4.3.0). **B** Segmentation of nuclei of the cells
1210 expressing the marker genes for; **i)** protonephridia and tegument (multiplexed) show that 6

1211 cells express Smp_335600 (*ShKT-domain protein*, a marker for protonephridia), 46 cells
1212 express *Meg-6* (a marker for tegument) (segmentation of 1 larva), and **ii)** seven cells that
1213 express an uncharacterised gene, Smp_318890 (a marker for parenchymal
1214 cells)(segmentation of 1 larva). **C & D)** The tegument marker *Meg-6* shows expression
1215 around nuclei in the posterior two-thirds of the larvae. The nuclei are below the body wall
1216 muscle, and cytoplasmic protrusions reach between muscle filaments (arrows) and form the
1217 epidermal ridges (arrowheads) between the ciliary plates (which are visible in **Dii & iii**
1218 expressing the ciliary plate marker Smp_096390) (asterisks). 100% of individuals examined,
1219 n > 30. **E)** The protonephridial marker Smp_335600 shows 'S'-shaped expression with
1220 transcripts extending from the nucleus of the anterior flame cell (AFC) along the excretory
1221 tubule and its nucleus (ETN) to the posterior flame cell (PFC), **Eii)** it is expressed around the
1222 nuclei (arrow) rather than the barrels (arrowhead) of the flame cells. 100% of individuals
1223 examined, n > 30. **Fi)** The pan-parenchymal marker, Smp_318890, was expressed in 2
1224 anterior cells (one on either side of the brain) and 5-7 cells posterior to the brain. **Fii)** These
1225 cells have long cytoplasmic protrusions that reach between all the other cells, **Fiii)** including
1226 the *ago2-1*$^{+}$ stem cells. 100% of individuals examined, n > 30. Scale shown in B also applies
1227 to C, D, Ei and F.
1228
1229 **Figure 5. Two defined populations of stem cells cluster by sex. A)** Expression profiles of
1230 cell marker genes that are specific or enriched in the stem cell clusters. Genes specific to the
1231 W (female-specific) sex chromosome are highlighted in blue. Genes validated by ISH are
1232 marked in red. Gene identifiers shown in parenthesis but with "Smp_" prefix removed for
1233 brevity. Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0). **B i)**
1234 UMAP including all genes shows that there are two Delta/Phi and two Kappa stem clusters,
1235 **ii)** UMAP showing that removal of all genes specific to the W and Z sex chromosomes
1236 results in one Delta/Phi and one Kappa cluster, indicating that the stem cells are
1237 transcriptionally different in male and female miracidia due to the expression of sex-linked
1238 genes. **C&D)** Multiplexed ISH showing three stem cell markers simultaneously in the same
1239 individual: *ago2-1* (Smp_179320) (pan-stem), *p53* (Smp_139530) (Delta/Phi) and *Uridine*
1240 *phosphorylase A* (*UPPA*, Smp_308140) (Kappa). **Ci)** *Ago2-1* expression reveals stem cells
1241 lateral and posterior to the brain (Optical section). 100% of individuals examined, n > 30. **ii)**
1242 MIP with segmentation of the 23 *ago2-1*$^{+}$ cells. **Di)** *P53* and *UPPA* expression shows the
1243 two stem cell populations intermingled in 100% of individuals examined, n > 30 **ii)** MIP and
1244 segmentation reveal there are more *p53*$^{+}$ cells than *UPPA*$^{+}$ cells (in this larva,15 are *p53*$^{+}$ and

1245 9 are *UPPA*$^{+}$). Scale shown in C also applies to D.

1246

1247 **Figure 6. RNA velocity analysis of stem and tegument cells from the miracidium and**
1248 **sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage origin
1249 of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field
1250 show the generalised direction of RNA velocity. **C)** Latent time analysis shows an estimated
1251 temporal-relationship between cells. **B)** and **C)** are based on the expression of both spliced
1252 and unspliced transcripts, and their expression dynamics across cells and genes. The phase
1253 plot, velocity, and expression were calculated for **D)** *p53-1*, **E)** *Zfp-1*, **F)** *p-53-2* and **G)** *eled*
1254 (no batch correction). Gene expression has been normalised (CPM) and log-transformed
1255 using scvelo(v. 0.2.4) For all of **D) - G)**, the phase plot (left plot) shows the proportion of
1256 spliced and unspliced transcripts in each cell, where each point is a cell and is coloured by the
1257 clusters in **B)**. The purple almond-shape overlaid represents the processes of transcription,
1258 splicing, and degradation, where this can be modelled. The dashed line shows the estimated
1259 steady state where RNA transcription is constant. The middle panel shows the RNA velocity,
1260 which for each gene is based on how the observations deviated from the estimated steady
1261 state towards induction or repression. The right panel shows gene expression. *p53-1* and *zfp-1*
1262 are predominantly expressed in the Delta/Phi-like miracidia stem cells and some miracidia
1263 tegument, and velocity indicates active expression of *p53-1* but downregulation of *zfp-1* in
1264 the stem cells. *p53-2* is most highly expressed in the Kappa-like miracidia stem cells, and
1265 velocity indicates this gene is being actively transcribed. *eled* expression is very low and only
1266 spliced transcripts have been detected, but is generally restricted to the Kappa-like miracidia
1267 stem cells.

1268 **Figure 7**. **A model for the fate of the two stem populations in the miracidium**. Adding
1269 single-cell data for miracidia stem cells (this study) to existing stem cell scenarios on
1270 *Schistosoma mansoni* development (Wang *et al*., 2018; Li *et al*., 2021) shows the continuum
1271 of the Kappa (κ) population from the miracidium, through the intra-molluscan stages to the
1272 juvenile and adult stages inside the mammalian host. Wang *et al*. (2018) proposed that the κ
1273 cells give rise to epsilon (*ε*), *eled*$^{+}$, cells in the juvenile primordial testes, ovaries, and
1274 vitellaria (germline), as well as in a gradient increasing toward the posterior growth zone
1275 (soma). They suggested that germ cells may be derived from ε-cells early in juvenile
1276 development, and *eled* is the earliest germline marker yet identified in schistosomes. This led
1277 to the idea that *S. mansoni* does not specify its germline until the juvenile stage (Wang *et al*.,

1278 2018) and a germline-specific regulatory program (including *eled, oc-1, akkn, nanos-1,*
1279 *boule*) was identified in intra-mammalian stages (Wang *et al*., 2018; Li *et al*., 2021). We
1280 show expression of these genes in κ stem cells in the miracidium. This suggests that after ~6
1281 days of embryogenesis, at hatching of the miracidium, the cells that may contribute to the
1282 germline might already be segregated into the κ population and the molecular regulatory
1283 program that differentiates somatic (delta/phi, $\delta/_{\varphi}$) and germ cell (κ) lineages is present.
1284 Furthermore, as *p53-2* plays a genotoxic stress response role in adult reproductive cells
1285 (Wendt *et al*., 2022), its expression in κ cells in the miracidia is another line of evidence that
1286 indicates that this population may contain the pluripotent stem cells that likely give rise to the
1287 germline.
1288

1289 **Figure 8**. **Tissue level segmentation of a miracidium reveals the contribution of each**
1290 **tissue to this simple larva**. *In situ* hybridisation using HCR (hybridisation chain reaction)
1291 enabled multiple tissue-level marker genes to be visualised simultaneously in the same larva.
1292 The nuclei of the cells of each tissue type were manually segmented using TrakEM2 in
1293 ImageJ (Cardona *et al*., 2012).
1294

1295 **Supplementary figure legends**
1296

1297 **Supplementary Figure 1. Contributions of sorted and unsorted cells to Seurat clusters.**
1298 This figure uses the same UMAP topology as presented in Figure 1C. **A)** Shows contribution
1299 of sorted cells to labelled clusters, **B)** shows contribution of unsorted cells to labelled
1300 clusters. **C)** UMAP where cells are coloured by sorted and unsorted treatment. **D)** UMAP
1301 where cells are coloured by sample of origin (samples 1 and 4 are sorted, samples 2 and 3 are
1302 unsorted).
1303

1304 **Supplementary Figure 2. Subclustering the Neuron 1 cluster using SAM revealed**
1305 **multiple distinct subpopulations**. **A)** The Neuron 1 cluster was analysed using Self-
1306 Assembling Manifold (SAM) algorithm, and Leiden clustering indicated 16 subpopulations.
1307 Library size was normalised and gene expression values were log-normalised using SAM
1308 (v1.0.1) and Scanpy (v1.8.2). **B)** The top five marker genes for each of the subpopulations.
1309 **C)** Expression of NPP PWamide, the top marker for subcluster 14. **D)** Expression of
1310 neuropeptide Y, the top marker for subcluster 13. **E)** Expression of neuropeptide F, the

1311 second top marker for subcluster 8. Although this cluster was difficult to define, this analysis
1312 suggests that there may be several small neural subpopulations within this cluster.
1313
1314 **Supplementary Figure 3. Neurons. A & B)** Top markers for cell cluster Neuron 1 include
1315 the neuropeptides precursor genes for Neuropeptide F, PWamide, and Neuropeptide Y. ISH
1316 of these genes shows expression in cells that make up the brain. **C)** G-alpha-q antibody
1317 staining reveals a girdle of 23 multi-ciliated nerve endings that sit between the 2nd and 3rd tier
1318 of ciliary plates. **D)** Neuron 4 marker Smp_319480 (putative mTORC1) transcripts are
1319 expressed at the base of these cilia; **E)** and some transcripts also reach the nuclei of circular
1320 body wall muscles. **F)** The expression of *p25 alpha* (which is a marker for ciliated cells in
1321 adult worms, Wendt *et al*., 2020) in the submuscular cells suggests that they may be ciliated.
1322

1323 **Supplementary Figure 4**. **Phylogeny of metazoan opsins.** There are four *Schistosoma*
1324 *mansoni* opsins that are expressed in two sub-clusters of Neuron 1; our analysis shows that
1325 three belong to the rhabdomeric opsin clade and the fourth is a peropsin. Scale bar shows
1326 branch lengths in expected amino acid substitutions per site. Shading of circles at internal
1327 nodes shows percentage support for the partition implied by that node from 1,000 non-
1328 parametric bootstrap samples.

1329

1330 **Supplementary Figure 5**. **The expression of two Polycystin genes and a Sox**
1331 **transcription factor in Neuron 4 cells. A)** Polycystin 1 and 2 are co-expressed in Neuron 4
1332 cells, orthologs in other animals form heterodimers in ciliated cells that function to detect
1333 flow sheer and vibrations (Kim *et al*., 2016; Calderon *et al*., 2018). Gene expression has been
1334 log-normalised and scaled using Seurat(v. 4.3.0). **B)** In planaria, a transcription factor Sox
1335 B1-2 regulates polycystin genes and together function in sensory neurons that detect flow.
1336 The co-expression of polycystin genes with a Sox transcription factor in Neuron 4 cells
1337 indicates that this transcription factor may regulate these polycystin genes in the miracidium.

1338

1339 **Supplementary Figure 6. STRING analysis of neural clusters and ciliary plates.**
1340 STRING analysis of the top 100 marker genes for each of the neural populations whose
1341 predicted networks have functional enrichment. **A)** Neuron 4, **B)** Neuron 5, **C)** Ciliary plates,

1342 **D)** Neuron 3. Neuron 1 is not a defined cluster and Neuron 2 showed no functional
1343 enrichment. Lines (edges) connecting nodes are based on evidence of the function of
1344 homologs. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein
1345 interaction). Minimum interaction (confidence) score of 0.7, corresponding to a high level of
1346 confidence, 0.4 medium level of confidence.
1347

1348 **Supplementary Figure 7. Gene expression by transcripts per million in the ciliary plates**
1349 **and gland cells.** Heatmap showing the gene expression, by transcripts per million, detected
1350 in the ciliary plates and gland cells hand-picked for plate-based scRNA-seq. Genes included;
1351 the top 20 significant genes between gland cell and ciliary plates, two neural markers, five
1352 VAL genes, five significant genes between gland cells and other cells, eight genes previously
1353 reported to produce secreted proteins in the miracidia (Wang *et al*., 2016; Wu *et al*., 2006),
1354 five top marker genes from the ciliary plate cluster.

1355

1356 **Supplementary Figure 8. STRING analysis of tegument and protonephridia clusters.**
1357 STRING analysis of the top 100 marker genes for **A)** tegument, **B)** protonephridia. Lines
1358 (edges) connecting nodes are based on evidence of the function of homologs. Functional
1359 enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum
1360 interaction (confidence) score of 0.7, corresponding to a high level of confidence, 0.4
1361 medium level of confidence.
1362

1363 **Supplementary Figure 9. STRING analysis of parenchyma clusters.** STRING analysis of
1364 the top 100 marker genes for **A)** Parenchyma 1, **B)** Parenchyma 2. Lines (edges) connecting
1365 nodes are based on evidence of the function of homologs. Functional enrichment (*FDR*) as
1366 provided by STRING (PPI = predicted protein interaction). Minimum interaction
1367 (confidence) score of 0.7, corresponding to a high level of confidence, 0.4 medium level of
1368 confidence.

1369

1370 **Supplementary Figure 10. Gene ontology (GO) analysis of marker genes identified in**
1371 **the miracidia cell clusters.** GO analysis was performed on each cluster using genes with a
1372 minimum AUC score of 0.7. Enriched biological process (BP) GO terms are shown for each
1373 cluster, and have been filtered to only show terms supported by a minimum of two genes.

1374 Top panels (left to right) show parenchyma, muscle, and tegument clusters, middle panels
1375 (left to right) show ciliary plate, neuron, and protonephridia clusters, and the bottom panel
1376 shows stem clusters. Full list of GO terms available in Supplementary Table 7.
1377
1378 **Supplementary Figure 11. Expression of known stem cell markers in miracidia cells.**
1379 **A-F)** Three stem cell classes were identified in the mother sporocyst stage and were named
1380 based upon their respective markers: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*nanos-2*$^+$ and *fgfrA*$^+$,
1381 *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$and, *fgfrB*$^+$ and *hesl*$^+$), as well as pan-stem marker *ago2-*
1382 *1* (Wang *et al*., 2018; Safarti *et al*., 2021). In the miracidium, we identified seven stem cell
1383 clusters by expression of *ago2-1*: stem A and B are *klf*$^+$ and *nanos-2*$^+$ (i.e. Kappa-like), stem
1384 C and D are *nanos-2*$^+$, *fgfrA*$^+$, *fgfrB*$^+$, and *hesl*$^+$ (i.e. they resembled both Delta and Phi).
1385 Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0).
1386
1387 **Supplementary Figure 12 - The effect of sex-linked genes on cell clustering.**
1388 *Schistosoma mansoni* has 7 pairs of autosomes and one pair of sex chromosomes; males have
1389 ZZ and females have ZW (Buddenborg *et al*., 2021). The sex chromosomes are composed of
1390 sex-specific regions that are unique to each chromosome flanked by pseudoautosomal regions
1391 that are common to both Z and W. The W-specific region (WSR) is a large, highly repetitive
1392 region that contains 38 genes, most of which are gametologs (genes with homologous copies
1393 on the WSR and ZSR). The Z-specific region (ZSR) contains 941 genes, of which only 35 are
1394 gametologs. **A & B)** The proportion of reads mapping to the WSR and ZSR across all cell
1395 type clusters shows **A)** putative male cells (Stem A and C i.e. those that lack WSR gene
1396 expression) and **B)** evidence of incomplete ZSR gene dosage compensation in the male stem
1397 cells. **C)** To determine whether any of the WSR genes were responsible for the clustering
1398 pattern, we excluded them from the mapping reference. The impact on the overall UMAP
1399 clustering was minimal. We repeated this excluding the **D)** ZSR genes, **E)** gametologues and
1400 **F)** ZSR genes, except the gametologues, in turn. In all four scenarios the stem cell clusters
1401 remained separated by sex. ZSR or WSR expression is therefore sufficient on their own to
1402 split these stem clusters in two.
1403
1404 **Supplementary Figure 13. Identification of new clusters after removing ZSR and WSR**
1405 **genes.** To identify whether stem cell clusters A and B, and C and D, still split down A/B and
1406 C/D lines despite the apparent collapsing of clusters after removing ZSR and WSR genes, we
1407 ran de novo clustering at several resolutions. Even once clusters were oversplitting, there was

1408 not a separation between stem cluster A/B and C/D. This was run using Seurat (v5.0.2), so
1409 the overall topology is not identical to that shown in Figure 5, though the same number of
1410 PCs are used. **A)** Shows the UMAP topology labelled by original cluster ID, **B)** shows the de
1411 novo clusters at resolution = 0.6, and **C)** shows the de novo clusters at resolution = 1.
1412
1413 **Supplementary Figure 14. Classification of cells based on percentage of reads mapping**
1414 **to W-specific region indicates incomplete dosage compensation of Z-specific region**
1415 **genes**. **A)** The percentage of reads mapping to W-specific genes by cluster. The blue line at
1416 0.2% indicates where the data were split into W+ and W- categories, i.e. Stem B and D
1417 contain cells with reads mapping to W-specific genes (and are likely cells from female
1418 miracidia), whereas most cells in Stem A and C do not have reads mapping to W-specific
1419 genes (cells from male miracidia). **B)** UMAP showing the clusters by W+ and W- categories.
1420 Almost complete separation can be seen in the main stem cell clusters, and the categories are
1421 mixed in most other clusters.
1422
1423 **Supplementary Figure 15. EdU labelling shows that there are no dividing cells during**
1424 **the free-swimming miracidia stage**. **A)** 4 hour EdU pulse on miracidia 0-6 hours post-
1425 hatching shows no EdU incorporation in any cells. **B)** Negative control. **C)** Positive control,
1426 three day EdU pulse on freshly-transformed mother sporocysts shows many dividing cells
1427 incorporating EdU, all treatments n = 30.
1428
1429 **Supplementary Figure 16. STRING analysis of the top 100 marker genes for the two**
1430 **stem cell populations in the miracidium (combining the female and male cells). A)** 61 of
1431 the top markers of Stem C and D combined (i.e. Delta/Phi-like) and **B)** 37 for Stem A and B
1432 combined (i.e. Kappa-like) form large predicted interactions enriched for translation,
1433 ribosome biogenesis and nucleic acid binding. Lines (edges) connecting nodes are based on
1434 evidence of the function of homologues. Functional enrichment (*FDR*) as provided by
1435 STRING. PPI = predicted protein interaction. Minimum interaction (confidence) score of 0.7,
1436 corresponding to a high level of confidence.
1437
1438 **Supplementary Figure 17. Differential gene expression between the two stem cell**
1439 **populations, Delta/Phi and Kappa**. **A)** Heatmap of top differentially expressed genes (by
1440 adjusted p-value) from each of Delta/Phi and Kappa populations show a high number of
1441 transcription factors (bold) in Delta/Phi and enrichment of lipid and glycolytic metabolism

genes in Kappa. Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0). **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 18. 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.* Gene expression has been normalised (CPM) and log-transformed using scvelo(v. 0.2.4). *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 19. 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. Gene expression has been normalised (CPM) and log-transformed using scvelo(v. 0.2.4).

Supplementary Figure 20. 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

1476 as calculated for genes that are mentioned in the text: red outline = known molecular
1477 regulators of germline stem cells in adult *S. mansoni*. Blue outline = pluripotency gene. *Oc-1*
1478 expression is low with little velocity signal, but the expression is predominantly seen in the
1479 Kappa-like miracidia stem cells (and Stem G). *Irx* is also primarily expressed in the Kappa-
1480 like miracidia stem cells with velocity results indicating an accumulation of spliced
1481 transcripts. *Irx* transcripts are also detected in Stem G and the sporocyst stem/germinal cells.
1482 *Akkn* expression is low and largely seen in the Kappa-like and Stem G miracidia stem
1483 clusters. Lsm14-domain containing protein transcripts are most consistently expressed in the
1484 Kappa-like miracidia stem cells. *Klf* is most highly expressed in sporocyst stem/germinal
1485 cells, and the Kappa-like miracidia stem cluster. Gene expression has been normalised
1486 (CPM) and log-transformed using scvelo(v. 0.2.4).
1487
1488 **Supplementary Figure 21. RNA velocity analysis of all tissue types from the miracidium**
1489 **and sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage
1490 origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity
1491 analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis
1492 shows an estimated temporal-relationship between cells. Phase plot, velocity and expression
1493 as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst
1494 cells combined with miracidia sample 1 cells). *P53-1* is most highly expressed in the
1495 Delta/phi-like miracidia stem clusters and some of the miracidia tegument cells, and the
1496 velocity indicates active transcription. *Zfp-1* is expressed most highly in the same clusters,
1497 but velocity suggests that the majority of transcripts are spliced. *P53-2* expression is highest
1498 in the Kappa-like miracidia stem clusters, but the expression is detected in several other
1499 clusters, including muscle, neuron, and other stem clusters. Velocity indicates that this gene
1500 is being actively transcribed in the majority of these clusters. *Eled* expression is very low, but
1501 restricted to the Kappa-like miracidia stem clusters. Gene expression has been normalised
1502 (CPM) and log-transformed using scvelo(v. 0.2.4)
1503
1504 **Supplementary Figure 22. RNA velocity analysis of all tissue types from the miracidium**
1505 **and sporocyst shows the dynamics of described germline-related genes. A)** UMAP shows
1506 the life cycle stage origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)**
1507 RNA velocity analysis flow field shows the generalized direction of RNA velocity. Phase
1508 plot, velocity and expression as calculated for **D)** *oc-1*, **E)** *boule*, **F)** *irx,* **G)** *akkn,* and **H)** an
1509 FDF domain-containing protein (no batch correction, sporocyst cells combined with

1510 miracidia sample 1 cells.). *Oc-1* expression is low but generally specific to the two Kappa-
1511 like miracidia stem cell clusters, and Stem G. Only spliced *Boule* transcripts were detected,
1512 and with low expression, predominantly in the Stem B and Stem E clusters. *Irx* was most
1513 highly expressed in Kappa-like miracidia stem clusters, as well as Stem E, Stem G, and
1514 sporocyst stem/germinal cells. *Akkn* is also largely expressed in the Kappa-like miracidia
1515 stem clusters. The FDF domain-containing protein expression is low, but seen most in
1516 Kappa-like miracidia stem clusters, Stem G, and several neural clusters. Gene expression has
1517 been normalised (CPM) and log-transformed using scvelo(v. 0.2.4).
1518
1519
1520 **Supplementary Figure 23. RNAvelocity analysis of all tissue types from the miracidium**
1521 **and sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage
1522 origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity
1523 analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis
1524 shows an estimated temporal-relationship between cells. Phase plot, velocity and expression
1525 as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst
1526 cells combined with miracidia sample 4 cells). *p53-1* and *zfp-1* are predominantly in the
1527 Delta/Phi-like miracidia stem cell clusters (Stem C and D) and also in the miracidia
1528 tegument. Velocity indicates that *p53-1* is being induced in the stem cells and repressed in the
1529 tegument, while *zfp-1* shows a gradient of velocity in the stem cells and repression in the
1530 tegument. *P53-2* is most highly expressed in the Kappa-like miracidia stem cells (Stem A and
1531 B) and Stem F, with velocity results indicating active transcription. *eled* expression was very
1532 low, but primarily in the Kappa-like miracidia stem cells (particularly Stem A). Gene
1533 expression has been normalised (CPM) and log-transformed using scvelo(v. 0.2.4).
1534
1535 **References**
1536

1537 1. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-
1538 cell RNA-seq artifacts: a molecular atlas of kidney development. Development.
1539 2017;144: 3625–3632. doi:10.1242/dev.151142

1540 2. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. In: R
1541 package version 2.54.0 [Internet]. 2023. doi:10.18129/B9.bioc.topGO

1542 3. Ashton PD, Harrop R, Shah B, Wilson RA. The schistosome egg: development and
1543 secretions. Parasitology. 2001;122: 329–338. doi:10.1017/s0031182001007351

1544 4. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to
1545 transient cell states through dynamical modeling. Nat Biotechnol. 2020;38:
1546 1408–1414. doi:10.1038/s41587-020-0591-3

1547 5. Bezares-Calderón LA, Berger J, Jasek S, Verasztó C, Mendes S, Gühmann M, et al.
1548 Neural circuitry of a polycystin-mediated hydrodynamic startle response for
1549 predator avoidance. Elife. 2018;7. doi:10.7554/eLife.36262

1550 6. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
1551 quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

1552 7. Buddenborg SK, Lu Z, Sankaranarayan G, Doyle SR, Berriman M. The stage- and
1553 sex-specific transcriptome of the human parasite Schistosoma mansoni. Sci
1554 Data. 2023;10: 775. doi:10.1038/s41597-023-02674-2

1555 8. Buddenborg SK, Tracey A, Berger DJ, Lu Z, Doyle SR, Fu B, et al. Assembled
1556 chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the
1557 evolution of its ZW sex-determination system. bioRxiv. 2021;
1558 2021.08.13.456314-2021.08.13.456314. doi:10.1101/2021.08.13.456314

1559 9. Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, et
1560 al. TrakEM2 software for neural circuit reconstruction. PLoS One. 2012;7:
1561 e38011–e38011. doi:10.1371/journal.pone.0038011

1562 10. Chernin E. Interference with the capacity of Schistosoma mansoni miracidia to
1563 infect the molluscan host. J Parasitol. 1968;54: 509–516.

1564 11. Choi HMT, Calvert CR, Husain N, Huss D, Barsi JC, Deverman BE, et al.
1565 Mapping a multiplexed zoo of mRNA expression. Development. 2016;143:
1566 3632–3637. doi:10.1242/dev.140137

1567 12. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J,
1568 et al. Third-generation in situ hybridization chain reaction: multiplexed,

1569 quantitative, sensitive, versatile, robust. Development. 2018;145.
1570 doi:10.1242/dev.165753

1571 13. Collins JJ 3rd, King RS, Cogswell A, Williams DL, Newmark PA. An atlas for
1572 Schistosoma mansoni organs and life-cycle stages using cell type-specific
1573 markers and confocal microscopy. PLoS Negl Trop Dis. 2011;5: e1009–e1009.
1574 doi:10.1371/journal.pntd.0001009

1575 14. Costain AH, MacDonald AS, Smits HH. Schistosome Egg Migration:
1576 Mechanisms, Pathogenesis and Host Immune Responses. Front Immunol.
1577 2018;9: 3042. doi:10.3389/fimmu.2018.03042

1578 15. Diaz Soria CL, Attenborough T, Lu Z, Fontenla S, Graham J, Hall C, Thompson S,
1579 Andrews TGR, Rawlinson KA, Berriman M, Rinaldi G. Single-cell
1580 transcriptomics of the human parasite Schistosoma mansoni first intra-
1581 molluscan stage reveals tentative tegumental and stem-cell regulators. Sci Rep.
1582 2024 Mar 12;14(1):5974. doi: 10.1038/s41598-024-55790-3. PMID: 38472267;
1583 PMCID: PMC10933418.

1584 16. Diaz Soria CL, Lee J, Chong T, Coghlan A, Tracey A, Young MD, et al. Single-
1585 cell atlas of the first intra-mammalian developmental stage of the human
1586 parasite Schistosoma mansoni. Nat Commun. 2020;11: 6411.
1587 doi:10.1038/s41467-020-20092-5

1588 17. Fishman L, Nechooshtan G, Erhard F, Regev A, Farrell JA, Rabani M. Single-cell
1589 temporal dynamics reveals the relative contributions of transcription and
1590 degradation to cell-type specific gene expression in zebrafish embryos. bioRxiv
1591 [Preprint]. 2023 Apr 21:2023.04.20.537620. doi: 10.1101/2023.04.20.537620.
1592 PMID: 37131717; PMCID: PMC10153228.

1593 18. Gąsiorowski L, Andrikou C, Janssen R, Bump P, Budd GE, Lowe CJ, et al.
1594 Molecular evidence for a single origin of ultrafiltration-based excretory organs.
1595 Curr Biol. 2021;31: 3629-3638.e2. doi:10.1016/j.cub.2021.05.057

1596 19. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al.
1597 Integrated analysis of multimodal single-cell data. Cell. 2021;184: 3573-
1598 3587.e29. doi:10.1016/j.cell.2021.04.048

1599 20. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016:
1600 expanding to enable helminth genomic research. Nucleic Acids Res. 2016;44:
1601 D774-80. doi:10.1093/nar/gkv1217

1602 21. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a
1603 comprehensive resource for helminth genomics. Mol Biochem Parasitol.
1604 2017;215: 2–10. doi:10.1016/j.molbiopara.2016.11.005

1605 22. IHME. 2019 Global Burden of Disease (GBD) study [Internet]. 2019. Available:
1606 https://vizhub.healthdata.org/gbd-results?params=gbd-api-2019-
1607 permalink/d30124bc2eaad3d977583362d4d37591 [vizhub.healthdata.org]

1608 23. Jurberg AD, Gonçalves T, Costa TA, de Mattos ACA, Pascarelli BM, de Manso
1609 PPA, et al. The embryonic development of Schistosoma mansoni eggs: proposal
1610 for a new staging system. Dev Genes Evol. 2009;219: 219–234.
1611 doi:10.1007/s00427-009-0285-9

1612 24. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang T-H, Liu Y, et al. Notum
1613 deacylates Wnt proteins to suppress signalling activity. Nature. 2015;519: 187–
1614 192. doi:10.1038/nature14259

1615 25. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell
1616 differential expression analysis. Nat Methods. 2014;11: 740–742.
1617 doi:10.1038/nmeth.2967

1618 26. Kim S, Nie H, Nesin V, Tran U, Outeda P, Bai C-X, et al. The polycystin complex
1619 mediates Wnt/Ca2+ signalling. Nat Cell Biol. 2016;18: 752–764.
1620 doi:10.1038/ncb3363

1621 27. Kumar P, Tan Y, Cahan P. Understanding development and stem cells using single
1622 cell-based analyses of gene expression. Development. 2017;144: 17–32.
1623 doi:10.1242/dev.133058

1624 28. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al.
1625 RNA velocity of single cells. Nature. 2018;560: 494–498. doi:10.1038/s41586-
1626 018-0414-6

1627 29. Li P, Nanes Sarfati D, Xue Y, Yu X, Tarashansky AJ, Quake SR, et al. Single-cell
1628 analysis of Schistosoma mansoni identifies a conserved genetic program
1629 controlling germline stem cell fate. Nat Commun. 2021;12: 485.
1630 doi:10.1038/s41467-020-20794-w

1631 30. Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba
1632 A. Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem
1633 Cells that Become Activated upon Brain Injury. Cell Stem Cell. 2015;17: 329–
1634 340. doi:10.1016/j.stem.2015.07.002

1635 31. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC, et al.
1636 EmptyDrops: distinguishing cells from empty droplets in droplet-based single-
1637 cell RNA sequencing data. Genome Biol. 2019;20: 63. doi:10.1186/s13059-
1638 019-1662-y

1639 32. Mair GR, Maule AG, Day TA, Halton DW. A confocal microscopical study of the
1640 musculature of adult Schistosoma mansoni. Parasitology. 2000;121 (Pt 2: 163–
1641 170. doi:10.1017/s0031182099006174

1642 33. Maldonado JF, Acosta-Matienzo J. Biological Studies on the Miracidium of
1643 Schistosoma Mansoni. Am J Trop Med. Arlington VA, USA: The American
1644 Society of Tropical Medicine and Hygiene; 1948;s1-28: 645–657.
1645 doi:10.4269/ajtmh.1948.s1-28.645

1646 34. Mann VH, Morales ME, Rinaldi G, Brindley PJ. Culture for genetic manipulation
1647 of developmental stages of Schistosoma mansoni. Parasitology. 2010;137: 451–
1648 462. doi:10.1017/S0031182009991211

1649 35. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in
1650 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell
1651 Syst. 2019;8: 329-337.e4. doi:https://doi.org/10.1016/j.cels.2019.03.003
1652 [doi.org]

1653 36. McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou X-N.
1654 Schistosomiasis. Nat Rev Dis Prim. 2018;4: 13. doi:10.1038/s41572-018-0013-
1655 8

1656 37. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KH (Joseph), et al. Modular,
1657 efficient and constant-memory single-cell RNA-seq preprocessing. Nat
1658 Biotechnol. 2021;39: 813–818. doi:10.1038/s41587-021-00870-2

1659 38. Moses L LP. BUSpaRse: kallisto | bustools R utilities [Internet]. 2023.
1660 doi:10.18129/B9.bioc.BUSpaRse

1661 39. Nanes Sarfati D, Li P, Tarashansky AJ, Wang B. Single-cell deconstruction of
1662 stem-cell-driven schistosome development. Trends Parasitol. 2021;37: 790–802.
1663 doi:https://doi.org/10.1016/j.pt.2021.03.005 [doi.org]

1664 40. Orosz F, Ovádi J. TPPP orthologs are ciliary proteins. FEBS Lett. 2008;582: 3757–
1665 3764. doi:10.1016/j.febslet.2008.10.011

1666 41. Pan SC. The fine structure of the miracidium of Schistosoma mansoni. J Invertebr
1667 Pathol. 1980;36: 307–372. doi:https://doi.org/10.1016/0022-2011(80)90040-3
1668 [doi.org]

1669 42. Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body
1670 axis. Cell. 2009;139: 1056–1068. doi:10.1016/j.cell.2009.11.035

1671 43. Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, et al. Dosage
1672 Compensation throughout the Schistosoma mansoni Lifecycle: Specific
1673 Chromatin Landscape of the Z Chromosome. Genome Biol Evol. 2019;11:
1674 1909–1922. doi:10.1093/gbe/evz133

1675 44. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of
1676 RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14: 687–
1677 690. doi:10.1038/nmeth.4324

1678 45. Poteaux P, Gourbal B, Duval D. Time series analysis of tegument ultrastructure of
1679 in vitro transformed miracidium to mother sporocyst of the human parasite

1680 Schistosoma mansoni. Acta Trop. 2023 Apr;240:106840. doi:
1681 10.1016/j.actatropica.2023.106840. Epub 2023 Jan 18. PMID: 36681315.

1682 46. Scimone ML, Cote LE, Reddien PW. Orthogonal muscle fibres have different
1683 instructive roles in planarian regeneration. Nature. 2017;551: 623–628.
1684 doi:10.1038/nature24660

1685 47. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-Cell RNA-
1686 Seq with Waterfall Reveals Molecular Cascades underlying Adult
1687 Neurogenesis. Cell Stem Cell. 2015;17: 360–372.
1688 doi:https://doi.org/10.1016/j.stem.2015.07.013 [doi.org]

1689 48. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, et al.
1690 Stem cell differentiation trajectories in Hydra resolved at single-cell resolution.
1691 Science (80-). American Association for the Advancement of Science;
1692 2019;365: eaav9314–eaav9314. doi:10.1126/science.aav9314

1693 49. Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, Reischl R,
1694 Taudes E, Lassnig M, Technau U. Single-cell transcriptomics identifies
1695 conserved regulators of neuroglandular lineages,Cell Reports, 40, 12,2022,
1696 111370, https://doi.org/10.1016/j.celrep.2022.111370.

1697 50. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al.
1698 STRING v11: protein–protein association networks with increased coverage,
1699 supporting functional discovery in genome-wide experimental datasets. Nucleic
1700 Acids Res. 2019;47: D607–D613. doi:10.1093/nar/gky1131

1701 51. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al.
1702 Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined
1703 Factors. Cell. 2007;131: 861–872. doi:https://doi.org/10.1016/j.cell.2007.11.019
1704 [doi.org]

1705 52. Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, et al.
1706 Mapping single-cell atlases throughout Metazoa unravels cell type evolution.
1707 Shalek AK, Barkai N, editors. Elife. eLife Sciences Publications, Ltd; 2021;10:
1708 e66747–e66747. doi:10.7554/eLife.66747

53. 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

54. 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

55. 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

56. 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

57. 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]

58. 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]

59. 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

60. 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,

1738 editor. Elife. eLife Sciences Publications, Ltd; 2018;7: e35449–e35449.
1739 doi:10.7554/eLife.35449

1740 61. Wang T, Zhao M, Rotgans BA, Strong A, Liang D, Ni G, et al. Proteomic Analysis
1741 of the Schistosoma mansoni Miracidium. PLoS One. Public Library of Science;
1742 2016;11: e0147247–e0147247.

1743 62. Wendt GR, Collins JNR, Pei J, Pearson MS, Bennett HM, Loukas A, et al.
1744 Flatworm-specific transcriptional regulators promote the specification of
1745 tegumental progenitors in Schistosoma mansoni. Patel NH, editor. Elife. eLife
1746 Sciences Publications, Ltd; 2018;7: e33221–e33221. doi:10.7554/eLife.33221

1747 63. Wendt GR, Shiroor DA, Adler CE, Collins JJ. Convergent evolution of a genotoxic
1748 stress response in a parasite-specific p53 homolog. Proc Natl Acad Sci.
1749 Proceedings of the National Academy of Sciences; 2022;119: e2205201119–
1750 e2205201119. doi:10.1073/pnas.2205201119

1751 64. Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, et al. A single-
1752 cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood
1753 feeding. Science (80-). 2020;369. doi:10.1126/science.abb7709

1754 65. WHO. Schistosomiasis [Internet]. 2023. Available: https://www.who.int/news-
1755 room/fact-sheets/detail/schistosomiasis [who.int]

1756 66. Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW. Muscle Cells
1757 Provide Instructions for Planarian Regeneration. Cell Rep. 2013;4: 633–641.
1758 doi:https://doi.org/10.1016/j.celrep.2013.07.022 [doi.org]

1759 67. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression
1760 data analysis. Genome Biol. 2018;19: 15. doi:10.1186/s13059-017-1382-0

1761 68. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al.
1762 Massively parallel digital transcriptional profiling of single cells. Nat Commun.
1763 2017;8: 14049. doi:10.1038/ncomms14049

1764

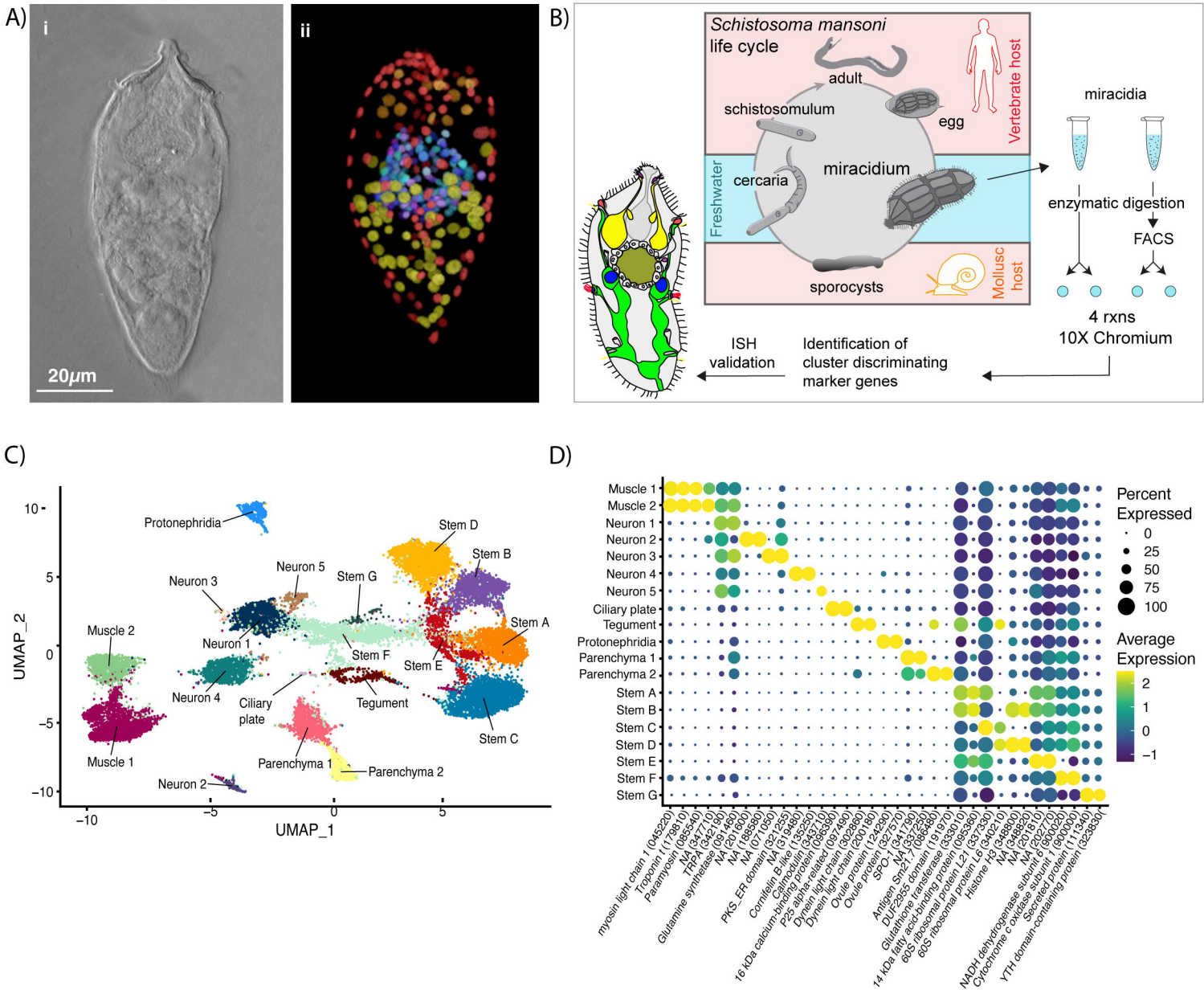
1765

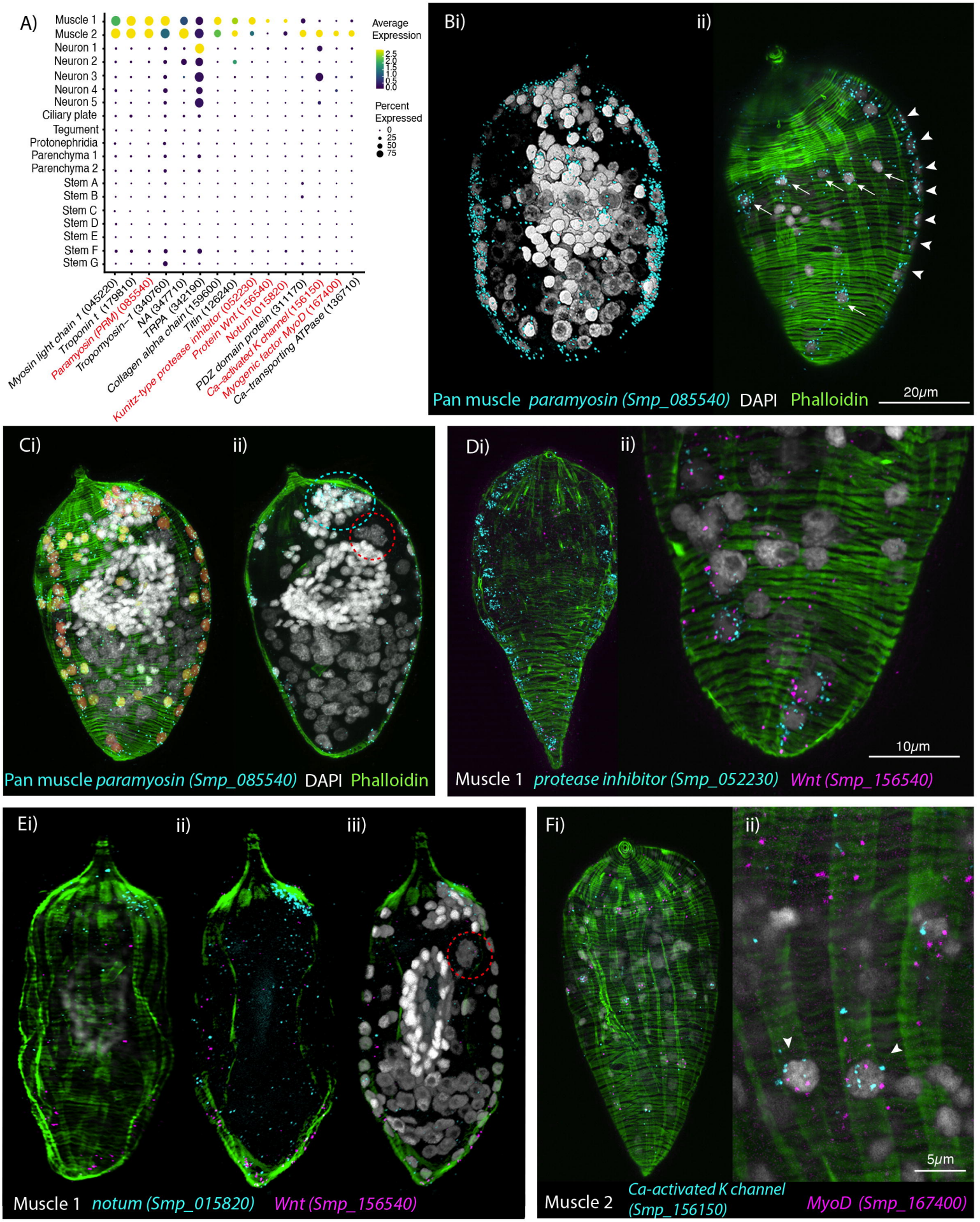
1766
1767
1768 **Supplementary Video legends**
1769
1770 **Supplementary Video 1** – Confocal z-stack of wholemount *in situ* hybridisation (WISH) of
1771 pan-muscle marker *paramyosin* (*PRM* Smp_085540) (cyan), counterstained with phalloidin
1772 (green) and DAPI (white), reveals which *PRM*$^{+}$ nuclei belong to the circular and longitudinal
1773 body wall muscles.
1774
1775 **Supplementary Video 2 –** Segmentation of *paramyosin* (Smp_085540) (cyan) positive cells
1776 (using the z-stack from Supplementary Figure 2); the nuclei of the circular body wall muscles
1777 are highlighted in orange, the nuclei of the longitudinal body wall muscles are highlighted in
1778 yellow.
1779
1780 **Supplementary Video 3 -** Confocal z-stack of WISH of muscle cluster 1 markers *Kunitz-*
1781 *type protease inhibitor* (Smp_052230) (cyan) and *Wnt-11-1* (Smp_156540) (magenta),
1782 counterstained with phalloidin (green) and DAPI (white), show expression in circular body
1783 wall muscles.
1784
1785 **Supplementary Video 4 –** Maximum intensity projection from confocal z-stack of WISH of
1786 muscle cluster 1 markers *notum* (Smp_015820) (cyan) and *Wnt-11-1* (Smp_156540)
1787 (magenta) counterstained with phalloidin (green) showing expression at opposite poles.
1788
1789 **Supplementary Video 5 -** Confocal z-stack of WISH of muscle cluster 2 markers
1790 *calcium activated potassium channel* (Smp_156150) (cyan) and *myoD* (Smp_167400)
1791 (magenta), counterstained with phalloidin (green) and DAPI (white), shows expression in
1792 longitudinal body wall muscles.
1793
1794 **Supplementary Video 6 -** Confocal z-stack of WISH of markers for neuron cluster 2
1795 (Smp_201600) (magenta) and neuron 3 (Smp_071050)(cyan), counterstained with DAPI
1796 (white), shows expression of both genes in cells whose nuclei form part of the nuclear rind of
1797 the brain.
1798

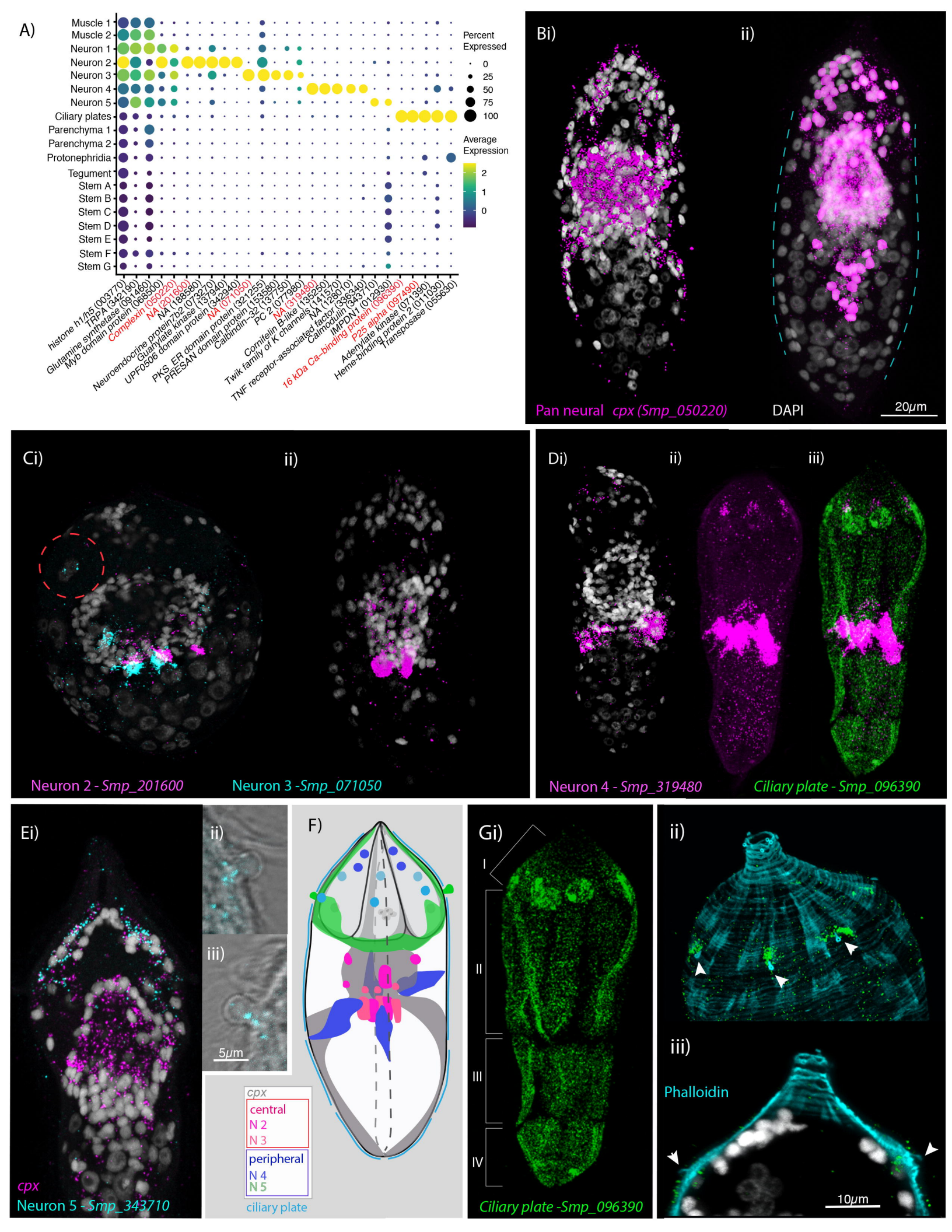
1799 **Supplementary Video 7 -** Confocal z-stack of WISH of ciliary plate marker *16 kDa calcium*
1800 *binding protein* (Smp_096390), counterstained with phalloidin (green) and DAPI (white),
1801 shows expression in the ciliary plates and six previously undescribed cells that we term sub-
1802 muscular cells. The nuclei of the sub-muscular cells sit beneath the body wall muscles, and
1803 each cell has a protrusion (possible cilia) that extends beyond the body wall muscles, between
1804 the first and second tiers of ciliary plates, and out to the exterior of the larva.
1805
1806 **Supplementary Video 8 –** Maximum intensity projection of confocal z-stack of WISH of
1807 ciliary plate marker *16 kDa calcium binding protein* (Smp_096390), counterstained with
1808 phalloidin (green) and DAPI (white), shows expression in ciliary plates and sub-muscular
1809 cells.
1810
1811 **Supplementary Video 9 –** Maximum intensity projection from confocal z-stack of WISH of
1812 neuron 2 marker Smp_319480 (magenta), counterstained with DAPI (white), showing
1813 expression in clusters of cells posterior to the brain that extend into the brain and out to the
1814 periphery.
1815
1816 **Supplementary Video 10 –** Maximum intensity projection from confocal z-stack of WISH
1817 of neuron 4 marker Smp_319480 (magenta) and ciliary plate marker *16 kDa calcium binding*
1818 *protein* (Smp_096390) (green), counterstained with DAPI (white), showing that the clusters
1819 of neuron 4 cells reach the periphery at the latitude between the second and third tiers of
1820 ciliary plates.
1821
1822 **Supplementary Video 11 -** Maximum intensity projection from confocal z-stack of WISH of
1823 neuron 5 marker Smp_319480 (cyan), counterstained with DAPI (white), showing expression
1824 in cells anterior to the brain and in and around the paired lateral glands and their secretory
1825 ducts.
1826
1827 **Supplementary Video 12 -** Maximum intensity projection from confocal z-stack of WISH of
1828 protonephridia marker *ShKT-domain containing protein* (Smp_335600) (magenta),
1829 counterstained with DAPI (white). This shows the transcripts in a bilaterally symmetrical
1830 pattern distributed along the s-shaped path of an excretory tubule that connects the anterior
1831 and posterior flame cells.
1832

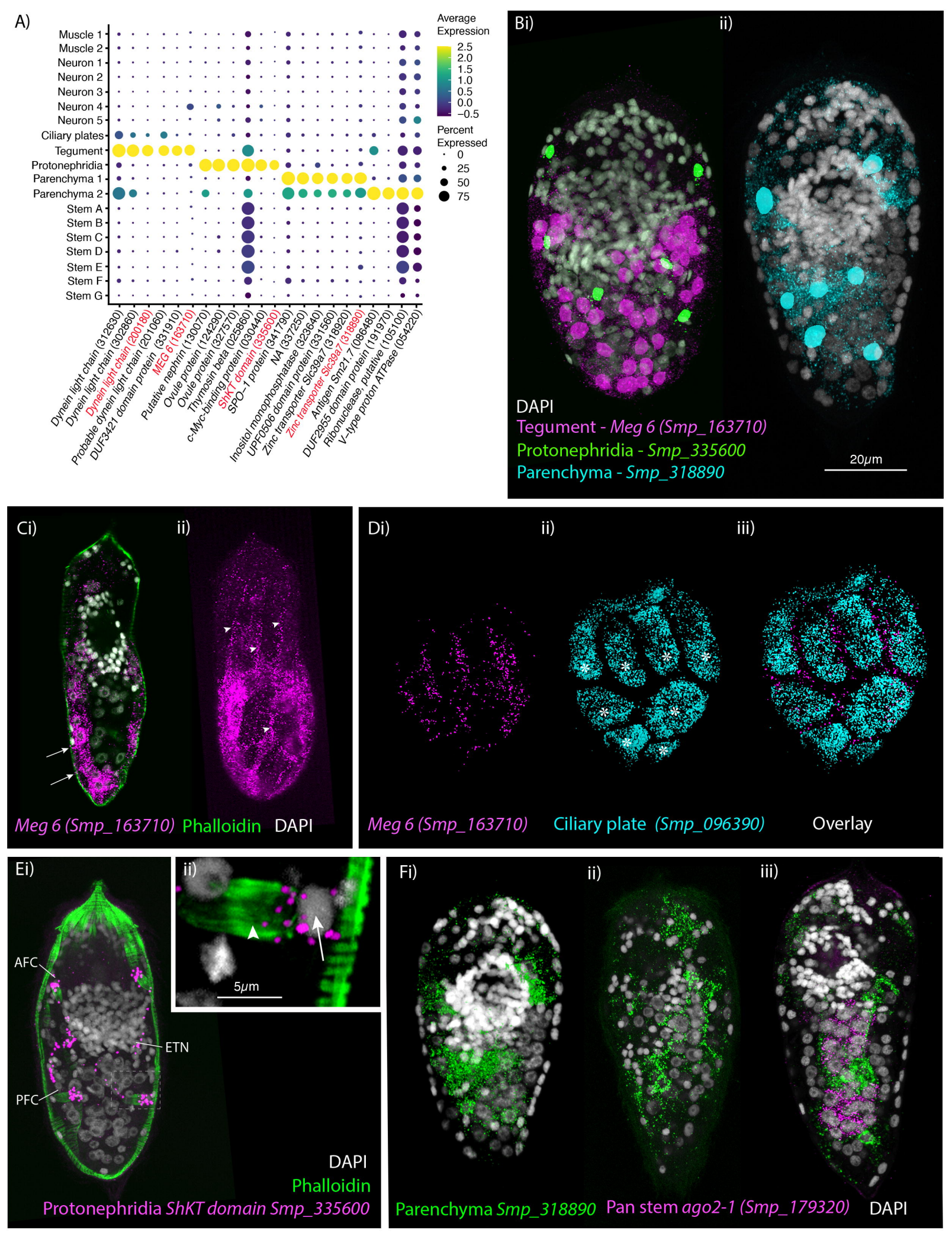
1833 **Supplementary Video 13 -** Confocal z-stack of WISH of markers for parenchyma 1 and 2
1834 (Smp_318890) (green) and stem cells *ago2-1* (Smp_179320) (magenta), counterstained with
1835 DAPI (white), shows the long cytoplasmic protrusions of the parenchymal cells reaching
1836 between the stem and other cells.
1837
1838 **Supplementary Video 14 -** Maximum intensity projection from confocal z-stack of WISH of
1839 marker for parenchyma 1 and 2 (Smp_318890) (green) counterstained with DAPI (white),
1840 shows the distribution of the parenchymal cells and their reach in the intercellular space of
1841 the larva.
1842
1843 **Supplementary Video 15 -** Confocal z-stack of WISH of markers for tegument *Meg6*
1844 (Smp_163710) (magenta) and *dynein light chain* (Smp_200180) (cyan), counterstained with
1845 phalloidin (green) and DAPI (white), shows the nuclei sit below the body wall muscle, and
1846 cytoplasmic protrusions reach between muscle filaments and form the epidermal ridges
1847 between the ciliary plates.
1848
1849 **Supplementary Video 16 -** Confocal z-stack of WISH of markers for stem clusters Delta/Phi
1850 *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (green) and stem E
1851 *Phosphomevalonate kinase* (Smp_072460) (magenta), counterstained with DAPI (white),
1852 shows there are no cells that are *pmk*-positive but negative for *UPPA* and *p53*. Stem E cells
1853 cannot be validated *in situ*.
1854
1855 **Supplementary Video 17 -** Confocal z-stack of WISH of markers for stem clusters Delta/Phi
1856 *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (magenta) and the pan-stem marker
1857 *ago2-1* (Smp_179320) (green), counterstained with DAPI (white), shows there are no cells
1858 that are *ago2-1* positive but negative for *UPPA* and *p53* that could be Stem F and G cells.
1859
1860 **Supplementary Video 18 -** Confocal z-stack of multiplexed WISH of tissue markers for
1861 stem *ago2-1* (Smp_179320) (green), muscle *paramyosin* (Smp_085540) (cyan) and nerve
1862 *complexin* (Smp_050220)(magenta).
1863
1864

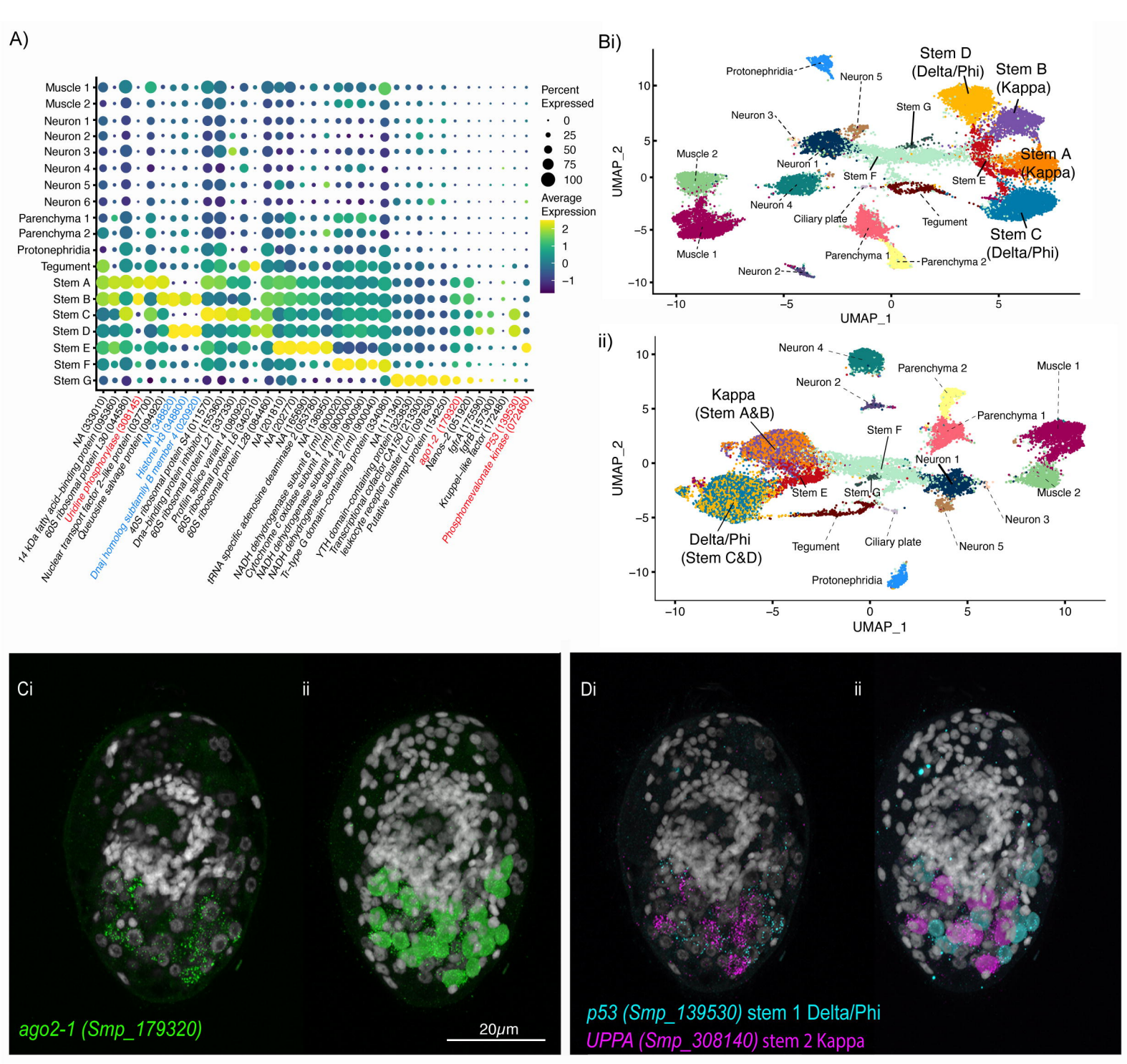
1865 **Supplementary Video 19 -** Confocal z-stack of multiplexed WISH of tissue markers for
1866 parenchyma (Smp_318890) (green), stem *ago2-1* (Smp_179320) (cyan) and tegument *Meg6*
1867 (Smp_163710) (magenta).
1868
1869 **Supplementary Video 20 -** Confocal z-stack of multiplexed WISH of tissue markers for
1870 tegument *Meg6* (Smp_163710) (green) and protonephridia *ShKT-domain containing protein*
1871 (Smp_335600) (magenta).
1872
1873
1874
1875
1876

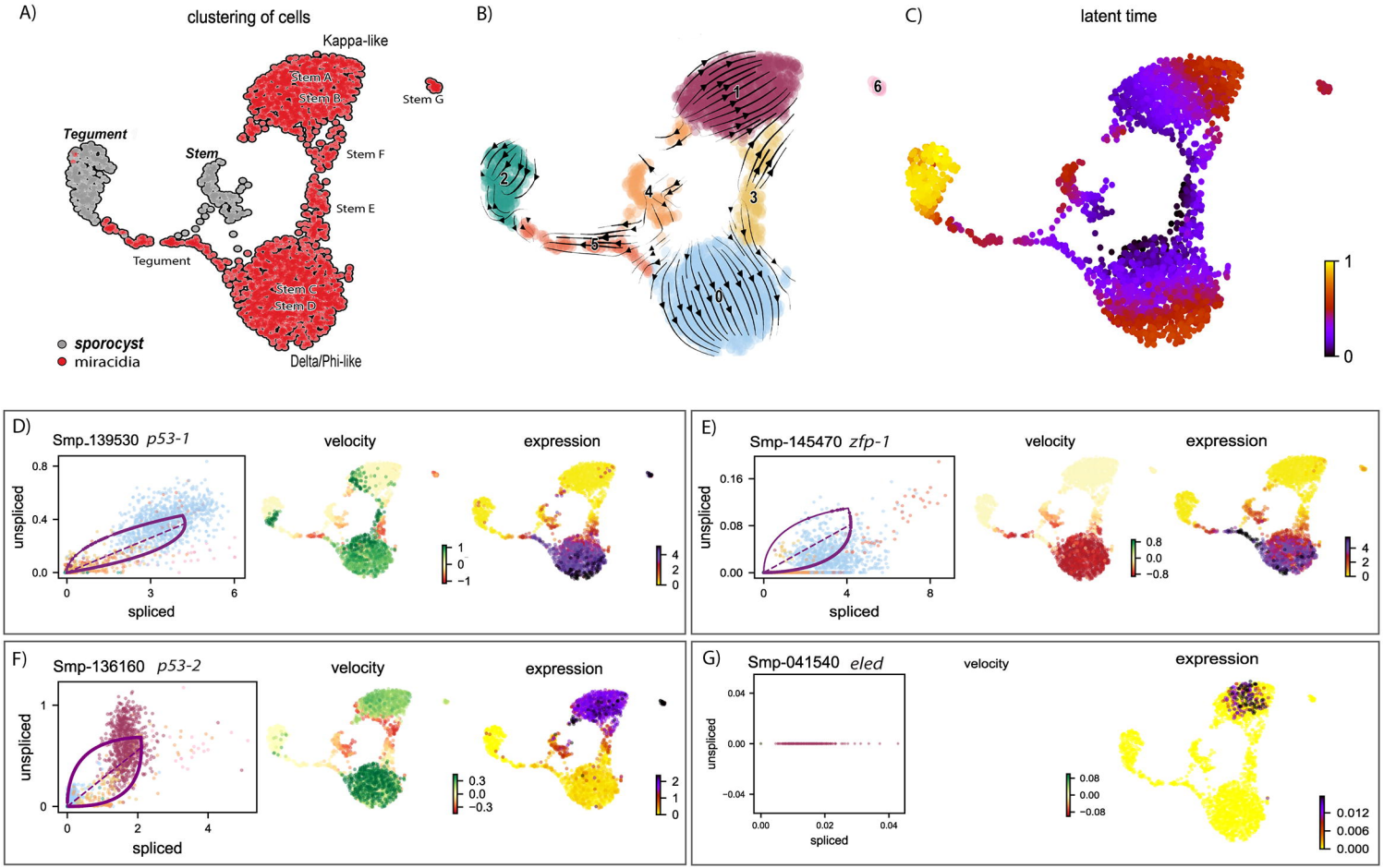

A)
i
ii
20μm
B)
Schistosoma mansoni
life cycle
adult
schistosomulum
egg
Vertebrate host
Freshwater
cercaria
miracidium
sporocysts
Mollusc host
miracidia
enzymatic digestion
FACS
4 rxns
10X Chromium
ISH
validation
Identification of
cluster discriminating
marker genes
C)
Protonephridia
Stem D
Stem B
Neuron 5
Neuron 3
Stem G
Stem A
Muscle 2
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
myosin light chain 1 (045220)
Troponin I (179810)
Paramyosin (085540)
NA (347710)
TRPA (342190)
Glutamine synthetase (091460)
NA (201600)
NA (188580)
NA (071050)
PKS_ER domain (321255)
NA (319480)
Cornifelin B-like (135250)
Calmodulin (343710)
16 kDa calcium-binding protein (096390)
P25 alpha-related (097490)
Dynein light chain (302860)
Dynein light chain (200180)
Ovule protein (124290)
Ovule protein (327570)
SPO-1 (341790)
NA (337250)
Antigen Sm217 (086480)
DUF2955 domain (191970)
Glutathione transferase (333010)
14 kDa fatty acid-binding protein (095360)
60S ribosomal protein L21 (337330)
60S ribosomal protein L6 (340210)
Histone H3 (348800)
NA (348820)
NA (201810)
NA (202770)
NADH dehydrogenase subunit 6 (900020)
Cytochrome c oxidase subunit 1 (900000)
Secreted protein (111340)
YTH domain-containing protein (323830)

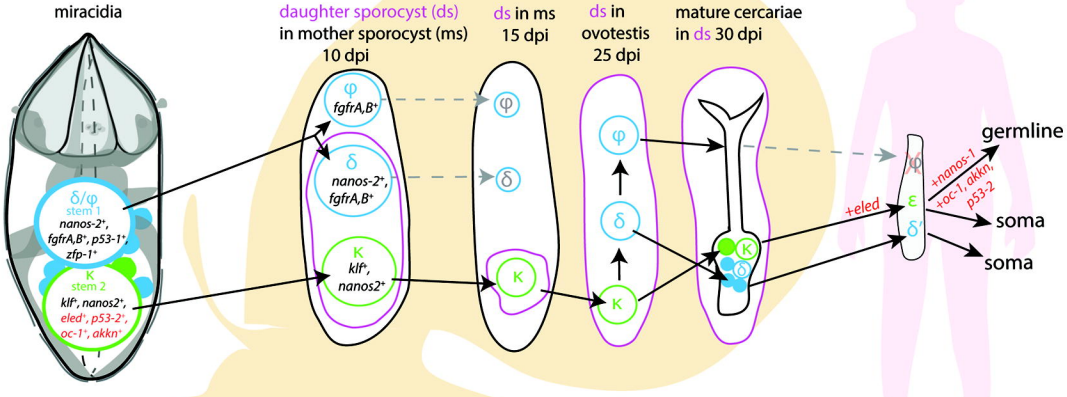

A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plate
Tegument
Protonephridia
Parenchyma 1
Parenchyma 2
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expressed
0
25
50
75
Myosin light chain 1 (045220)
Troponin t (179810)
Paramyosin (PRM) (085540)
Tropomyosin-1 (340760)
NA (347710)
TRPA (342190)
Collagen alpha chain (159600)
Titin (1262240)
Kunitz-type protease inhibitor (052230)
Protein Wnt (1565540)
Notum (015820)
PDZ domain protein (311170)
Ca-activated K channel (156150)
Myogenic factor MyoD (167400)
Ca-transporting ATPase (136710)
Bi)
ii)
Pan muscle paramyosin (Smp_085540) DAPI Phalloidin
20μm
Ci)
ii)
Pan muscle paramyosin (Smp_085540) DAPI Phalloidin
Di)
ii)
10μm
Muscle 1 protease inhibitor (Smp_052230) Wnt (Smp_156540)
Ei)
ii)
iii)
Muscle 1 notum (Smp_015820) Wnt (Smp_156540)
Fi)
ii)
5μm
Muscle 2 Ca-activated K channel (Smp_156150) MyoD (Smp_167400)

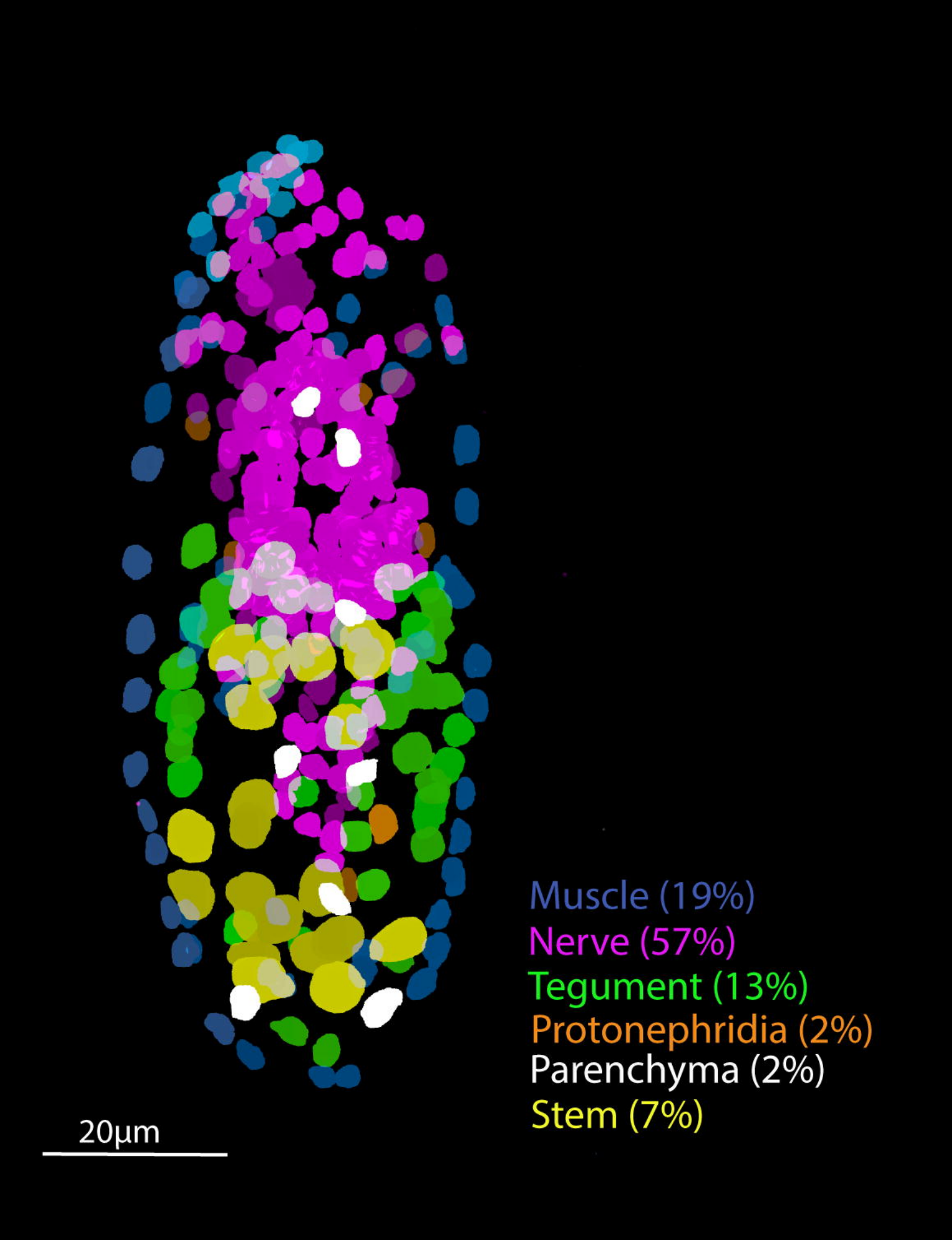

A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Ciliary plates
Parenchyma 1
Parenchyma 2
Protonephridia
Tegument
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
histone h1/h5 (003770)
TRPA (342190)
Glutamine synthetase (091460)
Myb domain protein (068500)
Complexin (050220)
NA (201600)
NA (188580)
Neuroendocrine protein 7b2 (073270)
Guanylate kinase (137940)
UPF0506 domain protein (342940)
PKS_ER domain protein (071050)
PRESAN domain protein (321255)
Calbindin-32 (153580)
PC 2 (077980)
NA (319480)
Cornifelin B-like (135250)
Twik family of K channels (141970)
TNF receptor-associated factor (128010)
Calmodulin (336340)
IMPDH1 (012930)
16 kDa Ca-binding protein (096390)
P25 alpha (097490)
Adenylate kinase (071390)
Heme-binding protein 2 (011030)
Transposase (055630)
Bi)
ii)
Pan neural cpx (Smp_050220)
DAPI
20μm
Ci)
ii)
Neuron 2 - Smp_201600
Neuron 3 -Smp_071050
Di)
ii)
iii)
Neuron 4 - Smp_319480
Ciliary plate - Smp_096390
Ei)
cpx
Neuron 5 - Smp_343710
ii)
iii)
5μm
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
10μm


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
Dynein light chain (312630)
Dynein light chain (302860)
Dynein light chain (200180)
Probable dynein light chain (201060)
DUF3421 domain protein (331910)
MEG 6 (163710)
Putative nephrin (130070)
Ovule protein (124290)
Ovule protein (327570)
Thymosin beta (025860)
c-Myc-binding protein (030440)
ShKT domain (335600)
SPO-1 protein (341790)
NA (337250)
Inositol monophosphatase (323640)
UPF0506 domain protein (331560)
Zinc transporter Slc39a7 (318920)
Zinc transporter Slc39a7 (318890)
Antigen Sm21.7 (086480)
DUF2955 domain protein (191970)
Ribonuclease, putative (105100)
V-type proton ATPase (054220)
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Percent Expressed
0
25
50
75
Bi)
ii)
DAPI
Tegument - Meg 6 (Smp_163710)
Protonephridia - Smp_335600
Parenchyma - Smp_318890
20μm
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
5μm
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 (339010)
14 kDa fatty acid–binding protein (095360)
60S ribosomal protein L30 (044580)
Uridine phosphorylase (308145)
Nucleat transport factor 2–like protein (037700)
Queuosine salvage protein (084920)
NA (348820)
Histone H3 (348800)
Dnaj homolog subfamily B member 4 (020920)
40S ribosomal protein S4 (011570)
Dna–binding protein inhibitor (153360)
Profilin splice variant 4 (090920)
60S ribosomal protein L21 (337330)
60S ribosomal protein L6 (340210)
60S ribosomal protein L28 (084460)
NA (201810)
NA (202770)
NA (165690)
tRNA specific adenosine deaminase 2 (053780)
NADH dehydrogenase subunit 6 (mt) (136850)
Cytochrome c oxidase subunit 1 (mt) (900020)
NADH dehydrogenase subunit 4 (mt) (900000)
NADH dehydrogenase subunit 2 (mt) (900090)
Ti–type G domain–containing protein (900040)
YTH domain–containing protein (334080)
NA (111140)
Transcriptional cofactor CA150 (323930)
Leukocyte receptor cluster (Lrc) (213300)
Putative unkempt protein (097830)
ago1-2 (179320)
Nanos–2 (154250)
IgfA (051920)
IgfB (175590)
Kruppel–like factor (175300)
P53 (139530)
Phosphomevalonate kinase (072480)
Bi)
Protonephridia
Neuron 5
Stem D
(Delta/Phi)
Stem B
(Kappa)
Neuron 3
Stem G
Muscle 2
Neuron 1
Stem F
Stem E
Stem A
(Kappa)
Neuron 4
Ciliary plate
Tegument
Stem C
(Delta/Phi)
Muscle 1
Parenchyma 1
Parenchyma 2
Neuron 2
UMAP_2
UMAP_1
10
5
0
−5
−10
ii)
Neuron 4
Parenchyma 2
Muscle 1
Neuron 2
Kappa
(Stem A&B)
Parenchyma 1
Stem F
Neuron 1
Stem E
Stem G
Muscle 2
Neuron 3
Delta/Phi
(Stem C&D)
Tegument
Ciliary plate
Neuron 5
Protonephridia
UMAP_2
UMAP_1
Ci
ii
ago2-1 (Smp_179320)
20μm
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
0
6
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+
K
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+
K
klf+,
nanos2+
ds in ms
15 dpi
φ
δ
K
ds in
ovotestis
25 dpi
φ
δ
K
mature cercariae
in ds 30 dpi
K
δ
+eled
φ
ε
δ
+nanos-1
+oc-1, akkn,
p53-2
germline
soma
soma


Muscle (19%)
Nerve (57%)
Tegument (13%)
Protonephridia (2%)
Parenchyma (2%)
Stem (7%)
20µm