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).

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 epidermal ridges of the tegument cells) and terebratorium (an apical modified epidermal plate with sensory structures and gland openings) (Pan, 1980).

101

102 To characterise this critical transmission stage, we have generated a single cell transcriptome
103 atlas for the miracidium. Our atlas provides molecular definitions for all described cell types
104 (Pan, 1980) and indicates the functions of different cell types within tissues. Multiplexed *in*
105 *situ* validation of cell type marker genes enabled the cell types to be spatially resolved within
106 the larva. Furthermore, the nuclear segmentation of the cells expressing tissue-specific genes
107 has revealed the relative contribution of each tissue type to the animal and each cell in the
108 larva has been assigned to a tissue. A key finding is the identification of two stem cell
109 populations in a miracidium, and advances in our understanding of the genome (Buddenborg
110 *et al*., 2021) has enabled us to determine that these two stem populations are transcriptionally
111 distinct between male and female larvae due to sex-linked gene expression. To predict the
112 fate of these two stem populations we carried out a trajectory analysis, combining scRNA-seq
113 data from miracidia and mother sporocysts, and this indicated that one stem cell population
114 likely contains the pluripotent cells that will develop into the cercariae, while the other
115 population gives rise to the somatic tissues, including the tegument. Similarities in gene
116 expression with the adult suggest that there is a tegument developmental program that is
117 redeployed at multiple stages of the complex life cycle.
118

119 **Results**

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

121 There are ~365 nuclei in a miracidium (median = 365, larva #1: 342, larva #2: 365, larva #3:
122 377). Some cell types, such as the apical gland and epithelial/tegument cells, are
123 multinucleate; and others, such as the ciliary plates may be anucleate (Pan, 1980, but cf
124 Meulemen *et al*., 1978). Therefore, the number of nuclei provides an approximation for the
125 actual number of cells (Figure 1A).
126

127 We performed scRNA-seq on a pool of ~20,000 mixed-sex miracidia, collected within four
128 hours of hatching from eggs (Figure 1B). Following dissociation, four samples of the cell
129 suspensions were collected; two were enriched for live cells using fluorescence-activated cell
130 sorting (FACS), and two were left unsorted to capture the maximum cellular diversity for the
131 atlas (Figure 1B). Using the droplet-based 10X Genomics Chromium platform,
132 transcriptome-sequencing data were generated from a total of 33,391 cells, of which 20,478
133 passed strict quality-control filters, resulting in a median of 858 genes and 1,761 median UMI
134 counts per cell (Supplementary Table 1). Given that an individual miracidium comprises

135 ~365 cells (Figure 1A), the number of quality-controlled cells theoretically represents >56-
136 fold coverage of each cell in the larva.
137
138 Using Seurat (version 4.3.0) (Hao *et al*., 2021), 19 distinct clusters of cells were identified,
139 along with putative marker genes best able to discriminate between the populations (Figure
140 1C & D and Supplementary Table 2 and 3). We used Seurat's JackStraw and ElbowPlot,
141 along with molecular cross-validation to select the number of principal components, and
142 Seurat's clustree to select a resolution where clusters were stable (Hao *et al*., 2021). Manually
143 curated lists of previously defined cell-type-specific genes from later developmental stages
144 (Diaz *et al*., 2020; Wendt *et al*., 2020; Nanes Sarfati *et al*., 2021) were compared against the
145 list of putative markers generated in this analysis. The expression of genes from both these
146 lists in the miracidia single-cell data was then examined to identify the cell types that each
147 Seurat cluster represented. Based on these markers, the 19 clusters were resolved into the
148 following cell populations: two muscle-like (3,934 cells), five resembling neurons (3,255
149 cells), one ciliary plate (40 cells), one tegumental (366 cells), one protonephridial (539 cells),
150 two parenchymal (1698 cells) and seven clusters resembling stem cells (10,686 cells) (Figure
151 1C).
152
153 There were contributions from sorted and unsorted samples in almost all clusters (except
154 ciliary plates). We found that some cell/tissue types had similar recovery from both methods
155 (e.g. Stem A, Muscle 2, and Tegument), others were preferentially recovered by sorting (e.g
156 Neuron 1, Neuron 4, and Stem E), and some were depleted by sorting (e.g. Parenchyma 1,
157 Protonephridia, and Ciliary plates) (Supplementary Figure 1), Supplementary Table 4). This
158 variation in recovery, therefore, enabled us to maximise the discovery and inclusion of
159 different cell types in the atlas.
160
161 **Orthogonal body wall muscle fibres are transcriptionally distinct**
162 The two muscle clusters were identified based on the expression of previously described
163 muscle-specific genes; *paramyosin* (Smp_085540), *troponin t* (Smp_179810) and *titin*
164 (Smp_126240) (Wendt *et al*., 2020; Diaz *et al*., 2020; Diaz & Attenborough *et al*., 2024), as
165 well as other differentially expressed markers; e.g. a putative collagen alpha chain
166 (Smp_159600) (Figure 2A). A crp1/csrp1/crip1 homolog (Smp_087250), encoding a
167 conserved transcriptional regulator of muscle (Tarashansky *et al*., 2021) was also expressed
168 in both muscle clusters. High expression of *paramyosin* (*PRM*; Smp_085540) was seen in

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

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

208 **Neural abundance and diversity**

209 We uncovered five clusters expressing the neural markers *Complexin* (*cpx*; Smp_050220),
210 *prohormone convertase 2* (*PC2*; Smp_077980) and Smp_068500 (a neural marker in adult
211 worms, Wendt *et al*., 2020) (Figure 3A). *Complexin* was expressed in and around at least 209
212 nuclei, indicating that 57% of cells in the miracidium were neurons; out of which 129 *cpx*$^{+}$
213 nuclei formed the nuclear rind of the brain (or neural mass/ring), and the remaining 80 were
214 situated peripherally, either anterior or posterior to the brain (Figure 3B). Sensory organelles
215 were identified for peripheral neural cell types. We were able to validate neuron clusters 2 - 5
216 using ISH of specific marker genes, and performed an *in silico* analysis of Neuron 1.
217
218 Neuron cluster 1 was large (1,470 cells), and expressed neuron-marker *complexin*, the
219 neuroendocrine precursor *7B2* (Smp_073270) and an uncharacterised gene (Smp_068500).
220 At this resolution, there were no clear marker genes restricted to Neuron 1. We explored
221 whether Neuron 1 could be further subdivided into transcriptionally distinct cells by
222 subclustering (Supplementary Figure 2; Supplementary Table 6) using the self-assembling
223 manifold (SAM) algorithm (Tarashansky *et al*., 2019) with ScanPy (Wolf *et al*., 2018), given
224 its reported strength in discerning subtle variation in gene expression (Tarashansky *et al*.,
225 2019), although a similar topology was subsequently found using Seurat. Three subclusters of
226 Neuron 1 expressed neuropeptide precursor genes amongst their top five markers; *PWamide*
227 (Smp_340860)(Wang *et al*., 2016) in subcluster 14, *Neuropeptide F* (Smp_088360) in
228 subcluster 8, and *Neuropeptide Y* (Smp_159950) in subcluster 13 (Supplementary Figure 2);
229 and each was expressed in the nuclear rind of the brain (Supplementary Figure 3). Another
230 subcluster was composed of putative photoreceptor cells (subcluster 7 - Supplementary
231 Figure 2, Supplementary Table 6) because their top-ten markers include two opsins
232 (Smp_180350 and Smp_332250) – identified here as a rhabdomeric opsin and a peropsin,
233 respectively (Supplementary Figure 4). Subcluster 7 also expressed other conserved
234 phototransduction cascade genes, such as a 1-phosphatidylinositol 4,5-bisphosphate

235 phosphodiesterase (Smp_245900) and a beta-arrestin (Smp_152280) (Supplementary Figure
236 2, Supplementary Table 6).
237
238 Neuron 2 expressed two further rhabdomeric opsins (Smp_180030 and Smp_104210;
239 Supplementary Figure 4), but across the cluster of 282 cells, the most expressed marker was
240 the uncharacterised gene Smp_201600. ISH revealed three strongly Smp_201600^{+} cells, with
241 nuclei in a median posterior location of the brain, and three cells with weaker expression in a
242 more anterior position (Figure 3C, Supplementary Video 6). Other top markers include *7B2,*
243 a guanylate kinase (Smp_137940) and a putative homeobox protein (Smp_126560).
244
245 Neuron 3 was a small cluster (47 cells) expressing *complexin* and *7B2*. The uncharacterised
246 gene, Smp_071050, was unique to and highly expressed throughout this population (Figure
247 3A). ISH showed expression in ~6 cells with projections that extended into the brain and to
248 the periphery, three of which sat adjacent to three Neuron 2 cells. There was also weak
249 expression in the four nuclei of the apical gland cell (Figure 3Ci, Supplementary Video 6).
250 STRING analysis of the top 100 markers of Neuron 3 predicted two protein interaction
251 networks with functional enrichment; these were associated with 'synapses' and 'calcium'
252 (Supplementary Figure 6). In Neuron clusters 1, 2 and 3, there was an enrichment for genes
253 annotated with Gene Ontology (GO) terms related to 'neuropeptide signalling pathway' and
254 'neurotransmitter transport' (Supplementary Table 7). There were, therefore, at least 5, and
255 likely several more, transcriptionally distinct cell types in the brain mass (we validated three
256 subclusters of Neuron 1, as well as Neurons 2 & 3).
257
258 Cells from neuron clusters 4 and 5 had nuclei peripheral to the brain. Neuron 4 formed a
259 large (1,140 cells), discrete cluster. The top marker was an uncharacterised gene,
260 Smp_319480 (Figure 3A), that, based on protein-structure prediction using I-TASSER (Yang
261 & Zhang, 2015), was similar to human mTORC1 (PDB entry 5H64; TM-score 0.85). *In situ*
262 expression revealed ~26 Smp_319480^{+} cells; four at the anterior end, with projections
263 leading to the body wall in the first tier of ciliary plates, and ~22 sat lateral and directly
264 posterior to the brain. These latter cells formed four clusters of ~3–7 cells, with Smp_319480
265 transcripts distributed around the nuclei and along projections leading into the brain and
266 extending out to the body wall terminating between the second and third tier of ciliary plates
267 (Figure 3D, Supplementary Video 9 and 10). Twenty-three multi-ciliated nerve endings
268 formed a girdle around the larva between the second and third tier of ciliary plates

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

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

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

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

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

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

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

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

402 *cysteine proteinase* (Smp_141610) and *serpin* (Smp_090080) (Diaz *et al.,* 2020; Wendt *et al.*,
403 2020; Diaz & Attenborough *et al*., 2024). All top marker genes for Parenchyma 1 were also
404 expressed in Parenchyma 2 but at lower levels, whereas the DUF2955 domain-containing
405 protein (Smp_191970) was uniquely expressed in Parenchyma 2 (Figure 4A). A gene
406 encoding a Zinc transporter Slc39a7 (Smp_318890) was a top marker for both clusters, with
407 expression confirmed *in situ* within 7–9 cells; two of which were located anteriorly on either
408 side of the brain and 5-7 located posterior to the brain (Figures 4B & F). These parenchyma
409 cells have long cytoplasmic protrusions that reach between all the other cells, filling the
410 intercellular space (Figure 4Fii-iii; Supplementary Videos 13 & 14). Pan (1980) called these
411 the interstitial cells and, observing abundant glycogen particles and lipid droplets, proposed
412 they function as sources of energy for the non-feeding larva and as nurse cells for the stem
413 cells during the first few days post-infection of the snail. We performed ISH for two
414 Parenchyma 2 markers, Smp_191970 and Smp_320330, but the first showed no detectable
415 signal, and the second had broad expression across all cells with no specific overlap with the
416 pan-parenchymal marker Smp_318890. Therefore, by ISH, we could only identify
417 parenchymal cells in general and could not identify the two clusters *in situ*. STRING analysis
418 for the Parenchyma 1 showed enrichment in networks for 'carbohydrate metabolic process'
419 and 'hydrolase', and for Parenchyma 2: 'phagosome', 'mTOR signalling pathway',
420 'mitochondrial inner membrane' and 'proton-transporting two sector atpase complex'
421 (Supplementary Figure 9). Analysis of GO annotations was able to distinguish between the
422 two parenchyma sub-populations, with Parenchyma 1 genes enriched for 'extracellular
423 region' and 'endopeptidase inhibitor activity', and Parenchyma 2 genes enriched for
424 processes that are 'dynein complex' and 'microtubule-based process' (Supplementary Table
425 7; Supplementary Figure 10).
426
427 **Two major stem cell populations in the miracidium that clustered by sex**
428 We identified stem cells based on the expression of known stem cell markers for the mother
429 sporocyst stage, e.g. *ago2-1* (Smp_179320), *nanos-2* (Smp_051920), *fgfrA* (Smp_175590),
430 *fgfrB* (Smp_157300), *klf* (Smp_172480) and *hesl* (Smp_024860)(Wang *et al*., 2013;
431 2018)(Figure 5A). The mother sporocyst has been reported to have three stem cell
432 populations: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*fgfrA*$^+$, *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$,
433 *fgfrB*$^+$ and *hesl*$^+$) based on scRNAseq of 35 stem cells (Wang *et al*., 2018). Wang *et al*. (2018)
434 suggest that the Kappa population serve as 'embryonic' stem cells that give rise to both the
435 Delta and Phi populations; Delta cells generate somatic tissues and Phi produce temporary

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

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

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

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

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

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

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

583

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

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

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

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

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

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

being actively transcribed, in the Kappa population. The incorporation of miracidia single-cell data into existing stem cell scenarios on *Schistosoma mansoni* development (Wang *et al*., 2013; Wang *et al*., 2018; Li *et al*., 2021) shows the continuum of the Kappa population from hatching through the intra-molluscan stages to the adult stage inside the mammalian host (Figure 7). The finding that *eled*, and many of the germline-specific regulatory program components, are expressed in some miracidia Kappa cells suggests that when a miracidium hatches, the cells that will contribute to the germline may already be segregated and the molecular regulatory program that differentiates somatic (Delta/Phi) and germ cell (Kappa) lineages is present. Intriguingly, the presence of spliced-only copies of the germline defining genes *eled* and *boule* could suggest that they are maternal transcripts that have been restricted to the primordial germ cells during embryogenesis, as is the case in Zebrafish embryos (Fishman et al., 2023). An alternative explanation is that unspliced transcripts exist for these lowly expressed genes but their abundance was below our threshold for detection.

An unexpected finding regarding the stem cells was the further clustering of the two populations by sex. This is the first identification of any sex differences in the miracidium. This degree of transcriptomic sexual difference was most evident in the Delta/Phi and Kappa stem cell clusters, with no distinct clustering by sex observed in other cell clusters. This could be explained by incomplete gene dosage compensation in the miracidium. Following investigations of differences in gene dosage compensation across developmental stages using bulk RNA-seq (Picard *et al*., 2019), our study suggests tissue-specific variation could add a further dimension to the role of dosage compensation in schistosomes. A similar phenomenon has been described in mouse embryonic stem cells and in *Drosophila,* where a precise regulation of a gene-by-gene dosage compensation mechanism is critical to the proper development of specific tissues and organs, such as the wing and eye (Valsecchi *et al*., 2018). In contrast to the miracidium, sex differences in undifferentiated stem cells have not been reported in later stages of the schistosome life cycle (Wang *et al.,* 2018; Diaz *et al.*, 2020; Wendt *et al*., 2020). In fact, it may not have been possible to detect these differences in earlier studies because scRNA-seq data were mapped onto older versions of the genome, where the sex chromosomes were not as well defined (Buddenborg *et al*., 2021). In addition to using an improved reference annotation, the present study has also benefited from a high level of coverage, enabling differences in stem cells to be more easily discerned. Whether these sex-linked gene expression and dosage differences have functional implications in developing male and female sporocysts remains to be investigated.

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

791

792 **Methods**

793 **Parasite material**

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

801

802 **Nuclei count**

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

837 0.5% BSA. The samples were centrifuged at 350 g for 10 minutes at 4°C, and all but 700 µl
838 of the supernatant was removed.
839
840 Sample 1 cell suspension was sorted; 200 µl was used for a no-stain control for
841 autofluorescence during cell counting and sorting. The remaining 500 µl of cell suspension
842 was stained with 1 µg/ml of Propidium Iodide (PI; Sigma P4864) to label dead/dying cells.
843 Sample 1 was sorted into an eppendorf tube using the BD Influx™ cell sorter (Becton
844 Dickinson, NJ) by enriching for PI negative cells (i.e., live cells). It took ~3 hours from the
845 beginning of the enzymatic digestion to generating single-cell suspensions ready for library
846 preparation on the 10X Genomics Chromium platform. Sample 2 cell suspension was not
847 sorted.
848
849 **10X Genomics library preparation and sequencing**
850 The 10X Genomics protocol ("Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User
851 Guide" available from
852 https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f
853 9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf) was followed
854 to create gel in emulsion beads (GEMs) containing single cells, hydrogel beads and
855 reagents for reverse transcription, perform barcoded cDNA synthesis, and produce
856 sequencing libraries from pooled cDNAs. The concentration of single-cell suspensions
857 was approximately 119 cells/µl (sample 1) and 122 cells/µl (sample 2), as estimated by flow
858 cytometry-based counting (Beckman Coulter Cytoflex S). We concentrated both samples to
859 732 cells/µl. We loaded two aliquots of sample 1 and two of sample 2 into four 10X reactions
860 according to the 10X protocol (Chromium Single Cell 3' Reagent Kits v3.1), intending to
861 capture approximately 10,000 cells per reaction.
862
863 Library construction (following GEM breakage) using 10X reagents following the "Single
864 Cell 3'Reagent Kits v3.1 User Guide". The four libraries were sequenced on one lane of
865 Illumina Novaseq XP (28 bp read 1, 90 bp read 2, 10 bp index 1, 10 bp index 2). All raw
866 sequence data will be deposited in the ENA.
867
868 **Mapping and quantification of single-cell RNA-seq**
869 We mapped the single-cell RNA-seq data to version 10 of the *S. mansoni* reference genome
870 (Buddenborg *et al*., (2021) and WormBase ParaSite ((Howe *et al*., 2016; 2017)) using 10X

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

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

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

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

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

1006 sequenced on one lane of NovaSeq SP. The reads from the 96 samples were pseudoaligned to
1007 the version 10 *S. mansoni* genome (WBPS18 release (Howe *et al*., 2016; 2017, with the same
1008 amendments as listed above) using kallisto (v 0.44.0) (Bray *et al*., 2016)). These were
1009 imported into R (v4.3.1) and the metadata were inspected, after which only samples coming
1010 from a single cell with a minimum of 20% reads pseudoaligned were retained (n=76). The
1011 samples were then processed using sleuth (v0.30.1) (Pimentel *et al*., 2017) following
1012 guidance from sleuth tutorials, including the steps for normalising, fitting error models, and
1013 differential analysis. Differential analysis was used to identify significant genes between the
1014 cell types, which were visualised, along with known genes of interest, using sleuth's heatmap
1015 function. Furthermore, the kallisto outputs were used to calculate the top 200 genes by
1016 median transcripts per million (TPM) of the two key cell types of interest, ciliary plates and
1017 gland cells, and these were used for GO term analysis following the method described below.
1018

1019 ***In situ* hybridisation**

1020 Miracidia collected as described above within 4 hours of hatching were fixed in 4% PFA in 2
1021 µm filtered PBSTw (1x PBS + 0.1% Tween) at room temperature for 30 minutes on a rocker.
1022 The larvae were transferred to incubation baskets (Intavis, 35 µm mesh) in a 24-well plate,
1023 rinsed 3 x 5 minutes in PBSTw, and incubated in 5 µg/ml Proteinase K (Invitrogen) in 1x
1024 PBSTw for 5 minutes at room temperature. They were post-fixed in 4% Formaldehyde in
1025 PBSTx for 10 minutes at room temperature and then rinsed in PBSTw for 10 minutes. From
1026 this point on, the *in situ* hybridisation experiments followed the protocol described by Choi *et*
1027 *al*., (2016; 2018) and developed for wholemount nematode larvae. We carried out fluorescent
1028 *in situ* hybridisation on 27 cell-type marker genes. Probes, buffers, and hairpins for third-
1029 generation *in situ* hybridization chain reaction (HCR) experiments were purchased from
1030 Molecular Instruments (Los Angeles, California, USA). Molecular Instruments designed
1031 probes against the sequences on WormBase ParaSite (WBPS version 18, WS271, Howe *et*
1032 *al*., 2017)(Supplementary Table 18). We subsequently labelled some larvae with phalloidin
1033 by adding 1 µl phalloidin to 1 ml 5X SSCT for an hour at room temperature in the dark, then
1034 rinsing 6 x 30 minutes in SSCT, incubating overnight at 4°C in DAPI fluoromount-G
1035 (Southern Biotech) and then mounting and imaging on a confocal laser microscope (Leica
1036 Sp8). We carried out at more than three *in situ* hybridisation experiments for each marker
1037 gene we validated (each experiment was a biological replicate). From each experiment we
1038 imaged (by confocal microscopy) 10 miracidia (technical replicates) per marker gene
1039 experiment. To count the number of cells belonging to each tissue type we imported confocal

1040 Z-stacks of tissue marker gene expression (Supplementary Videos 18-20) into ImageJ, and
1041 using the plugin TrakEM2 (Cardona *et al*., 2012), manually segmented each nuclei that was
1042 expressing, and was surrounded, by the transcripts.
1043

1044 **Opsin phylogeny**

1045

1046 Amino acid sequences for selected metazoan opsin proteins were obtained from the GenBank
1047 database and aligned using the —auto mode of MAFFT v7.450 (Katoh & Standley, 2013),
1048 then trimmed using trimAl v1.4.rev22 (Capella-Gutierrez *et al*., 2009) to remove alignment
1049 columns with more than 50% gap characters ensuring at least 50% of alignment columns
1050 were maintained in the final alignment. Phylogenetic analysis of this alignment was
1051 performed using RaxML v8.2.12 (Stamatakis, 2014). The best-fitting empirical amino acid
1052 substitution model was identified as maximising the log-likelihood of the data under a model
1053 incorporating gamma-distributed variation in substitution rates across sites, and this optimal
1054 model (LG) was used for subsequent inference. The maximum-likelihood tree and branch
1055 lengths were identified as the highest likelihood from 25 independent searches starting from
1056 different random starting tree topologies, and support for partitions on this tree was estimated
1057 using 1000 bootstrap replicates of resampling alignment positions with replacement, using
1058 the fast bootstrap (-x) algorithm implemented in RaxML.
1059

1060 **Protein-Protein interaction analysis**

1061 Protein interactions were predicted using the online search tool STRING (www.string-db.org;
1062 V 11) (Szklarczyk *et al.*, 2019) for the following well-defined cell clusters: Neuron 3-5,
1063 ciliary plates, protonephridia, Parenchyma 1 & 2, tegument and the two main stem cell
1064 populations, Delta/Phi and Kappa. Protein sequences for the top 100 marker genes for each
1065 cell cluster were collected from WormBase ParaSite. The protein sequences were entered as a
1066 multiple protein search. Default settings were used to predict interactions with either a
1067 minimum interaction (confidence) score of 0.4 or 0.7, corresponding to a medium or high
1068 level of confidence.
1069

1070 **Gene Ontology enrichment analysis**

1071 Each cluster's list of marker genes was filtered to focus on specific cluster markers. Genes
1072 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.

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

1106 TA and KR drafted the complete manuscript, with sections of text contributed by GR and
1107 MB
1108 JC carried out the opsin phylogeny
1109 All authors read and approved the final manuscript.
1110

1111 **Acknowledgements**
1112 We thank Simon Clare and his team for technical assistance with the life cycle maintenance
1113 and collection of parasite material at the Wellcome Sanger Institute, and Karl Hoffmann,
1114 Josephine Fforde-Thomas and Benjamin Hulme for additional ongoing parasite material. We
1115 thank Sarah Buddenborg for work on the annotation of the *Schistosoma mansoni* genome;
1116 David Goulding for assistance with Microscopy; teams within DNA Pipelines for data
1117 production; and core Informatics teams and Pathogen Informatics for the infrastructure used
1118 for data analysis. We thank members of the Parasite Genomics team at the Wellcome Sanger
1119 Institute for their comments and input on this study. Finally, we thank the reviewers and the
1120 editor for their valuable feedback on this manuscript.
1121
1122

1123 **Figure legends**
1124

1125 **Figure 1**. **Identification of 19 transcriptionally distinct cell types in the miracidium.**
1126 **A)** The miracidium is composed of ~365 cells; **i)** DIC image of miracidium, **ii)** a Maximum
1127 Intensity Projection (MIP) of a confocal z-stack of DAPI-stained miracidium with nuclei
1128 segmented to enable counting (larval anterior pole at the top in all images). **B)** Experimental
1129 scheme describing the parasite's life cycle (images of developmental stages not to scale),
1130 parasite dissociation, single-cell analysis of miracidium and validation pipeline. An average
1131 of 9975 miracidia per sample were dissociated; two samples were enriched for live cells
1132 (propidium iodide negative) using fluorescence-activated cell sorting (FACS), another two
1133 samples were unsorted. Cells were loaded according to the 10X Chromium single-cell 3'
1134 protocol. Clustering was carried out to identify distinct populations and population-specific
1135 markers. Validation of population-specific markers was performed by *in situ* hybridisation
1136 (ISH). **C)** Uniform Manifold Approximation and Projection (UMAP) representation of
1137 20,478 miracidium single cells. Cell clusters are coloured and distinctively labelled by cluster
1138 identity. **D)** Gene-expression profiles of the top population markers identified for each cell
1139 cluster (gene identifiers shown in parenthesis but with "Smp_" prefix removed for brevity).

1140 The colours represent the expression level from yellow (high expression) to dark blue (low
1141 expression). Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0). The
1142 sizes of the circles represent the percentages of cells in those clusters that expressed a specific
1143 gene.
1144
1145 **Figure 2. Orthogonal body wall muscles are transcriptionally distinct. A)** Dot plot
1146 depicting the expression profiles of specific- or enriched- marker genes in the two muscle
1147 clusters. Genes validated by *in situ* hybridisation (ISH) are marked in red. Gene identifiers
1148 shown in parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been
1149 log-normalised and scaled using Seurat(v. 4.3.0). **B) i-ii)** Wholemount ISH (WISH) of
1150 paramyosin *PRM*$^+$ and counterstaining with phalloidin reveals the nuclei of the circular body
1151 wall muscles (BWM), which form two distinct bilaterally symmetrical lines that run
1152 peripherally from the pole of the larva to the other (arrowheads in ii). The longitudinal BWM
1153 nuclei (ii - arrows) are spaced regularly between the lines of the circular BWM nuclei and
1154 their actin fibres run orthogonally to the circular muscles. These patterns were seen in 100%
1155 of individuals examined, n = 50. **C) i)** Segmentation of the *PRM*$^+$ cells in one miracidium
1156 shows 74 muscle cells in total: ~28 circular (segmented in orange), 33 longitudinal (yellow)
1157 and **ii)** 13 in an anterior unilateral cluster (inside dashed cyan line) adjacent to the apical
1158 gland cell (identifiable by its four nuclei, inside dashed red circle). **D)** Expression of markers
1159 for muscle 1 cluster; **i)** *Kunitz-type protease inhibitor* is expressed in circular BWM and *wnt-*
1160 *11-1* is expressed in 7 *Kunitz-type protease inhibitor*$^+$ cells at the posterior pole; **ii)** close-up
1161 of posterior end of another miracidium showing the co-expression of *Kunitz-type protease*
1162 *inhibitor* and *wnt-11-1.* In 100% of individuals examined, n = 30. **E) i-iii)** *Notum,* an inhibitor
1163 of *wnt* signalling, is expressed highly at the opposite pole to *wnt-11-1* in the cluster of 13
1164 muscle nuclei adjacent to the apical gland cell (nuclei in dashed red circle). In 100% of
1165 individuals examined, n = 30. **F)** Expression of markers for muscle 2 cluster; a *calcium*
1166 *activated potassium channel* (Smp_156150) and the transcription factor *myoD* (Smp_167400)
1167 show co-expression in longitudinal BWMs (arrowheads). In 100% of individuals examined, n
1168 = 30. Scale shown in B also applies to C, Di, E and Fi.
1169
1170
1171 **Figure 3. Neural complexity and ciliary plates in a simple larva. A)** Expression profiles of
1172 cell marker genes that are specific or enriched in the five neuronal clusters and ciliary plates.
1173 Genes validated by *in situ* hybridisation (ISH) are marked in red. Gene identifiers shown in

1174 parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been log-
1175 normalised and scaled using Seurat(v. 4.3.0). *Complexin* (*cpx*) is expressed in all five nerve
1176 clusters and is a neural marker in the adult (Wendt *et al*., 2020). **B) i)** A dorso-ventral view
1177 of WISH of *cpx* shows expression in the brain, to the anterior - with projections around the
1178 three gland cells, and in two posterior clusters (100% of individuals examined, n = 50). **ii)**
1179 Segmentation of *cpx*$^{+}$ cells (in magenta) in one miracidium (lateral view) reveals that at least
1180 209 of the 365 cells that make up the larva are neural and that the clusters of *cpx* cells that are
1181 located posterior to the brain, and send projections to the posterior pole, sit perpendicular to
1182 the circular muscle cell nuclei (blue dashed lines). **C)** Multiplexed ISH for top markers for
1183 neuron 2 (Smp_201600) and neuron 3 (Smp_071050) shows expression in multiple cells of
1184 the brain, including cells adjacent to each other. Smp_071050 is also expressed in the apical
1185 gland cell (red dashed circle). 100% of individuals examined, n = 30. **D)** The top marker for
1186 neuron 4 Smp_319480 is expressed in four cells anteriorly (arrows) and ~22 that sit lateral
1187 and just posterior to the brain; these later cells send projections into the brain and out to the
1188 body wall at the latitude between the 2nd and 3rd tier of ciliary plates **(iii)** seen here expressing
1189 the ciliary plate marker gene Smp_096390. 100% of individuals examined, n > 30. **E)** Neuron
1190 5 marker Smp_343710*,* an EF-hand domain-containing protein, is expressed in **(i)** 10-20 cells
1191 whose nuclei sit outside of, and anterior to, the brain, in and around the paired lateral glands
1192 and their secretory ducts, and **(Eii & iii, F)** in a pair of bilaterally symmetrical bulbous
1193 protrusions. 100% of individuals examined, n = 30. **F)** Summary schematic of the *in situ*
1194 expression of marker genes for the neural cell clusters. **G) i)** A top marker for ciliary plates, a
1195 calcium binding protein (Smp_096390), shows transcripts expressed in all the ciliary plates
1196 of all four tiers (brackets) and in 6 cells towards the anterior pole. **(ii & iii)** Counterstaining
1197 with phalloidin shows that the nuclei of these anterior cells sit beneath the body wall muscle
1198 and send a projection externally between the 1st and 2nd tier of ciliary plates (arrowheads).
1199 100% of individuals examined, n > 30. Scale shown in B also applies to C, D, Ei and Gi.
1200
1201
1202 **Figure 4. Identification of the tegument, protonephridia, parenchymal cells in the**
1203 **miracidium**. **A)** Expression profiles of cell marker genes specific or enriched in these cell
1204 type clusters. Genes validated by *in situ* hybridisation are in red. Gene identifiers shown in
1205 parenthesis but with "Smp_" prefix removed for brevity. Gene expression has been log-
1206 normalised and scaled using Seurat(v. 4.3.0). **B** Segmentation of nuclei of the cells
1207 expressing the marker genes for; **i)** protonephridia and tegument (multiplexed) show that 6

1208 cells express Smp_335600 (*ShKT-domain protein*, a marker for protonephridia), 46 cells
1209 express *Meg-6* (a marker for tegument) (segmentation of 1 larva), and **ii)** seven cells that
1210 express an uncharacterised gene, Smp_318890 (a marker for parenchymal
1211 cells)(segmentation of 1 larva). **C & D)** The tegument marker *Meg-6* shows expression
1212 around nuclei in the posterior two-thirds of the larvae. The nuclei are below the body wall
1213 muscle, and cytoplasmic protrusions reach between muscle filaments (arrows) and form the
1214 epidermal ridges (arrowheads) between the ciliary plates (which are visible in **Dii & iii**
1215 expressing the ciliary plate marker Smp_096390) (asterisks). 100% of individuals examined,
1216 n > 30. **E)** The protonephridial marker Smp_335600 shows 'S'-shaped expression with
1217 transcripts extending from the nucleus of the anterior flame cell (AFC) along the excretory
1218 tubule and its nucleus (ETN) to the posterior flame cell (PFC), **Eii)** it is expressed around the
1219 nuclei (arrow) rather than the barrels (arrowhead) of the flame cells. 100% of individuals
1220 examined, n > 30. **Fi)** The pan-parenchymal marker, Smp_318890, was expressed in 2
1221 anterior cells (one on either side of the brain) and 5-7 cells posterior to the brain. **Fii)** These
1222 cells have long cytoplasmic protrusions that reach between all the other cells, **Fiii)** including
1223 the *ago2-1*$^{+}$ stem cells. 100% of individuals examined, n > 30. Scale shown in B also applies
1224 to C, D, Ei and F.
1225
1226 **Figure 5. Two defined populations of stem cells cluster by sex. A)** Expression profiles of
1227 cell marker genes that are specific or enriched in the stem cell clusters. Genes specific to the
1228 W (female-specific) sex chromosome are highlighted in blue. Genes validated by ISH are
1229 marked in red. Gene identifiers shown in parenthesis but with "Smp_" prefix removed for
1230 brevity. Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0). **B i)**
1231 UMAP including all genes shows that there are two Delta/Phi and two Kappa stem clusters,
1232 **ii)** UMAP showing that removal of all genes specific to the W and Z sex chromosomes
1233 results in one Delta/Phi and one Kappa cluster, indicating that the stem cells are
1234 transcriptionally different in male and female miracidia due to the expression of sex-linked
1235 genes. **C&D)** Multiplexed ISH showing three stem cell markers simultaneously in the same
1236 individual: *ago2-1* (Smp_179320) (pan-stem), *p53* (Smp_139530) (Delta/Phi) and *Uridine*
1237 *phosphorylase A* (*UPPA*, Smp_308140) (Kappa). **Ci)** *Ago2-1* expression reveals stem cells
1238 lateral and posterior to the brain (Optical section). 100% of individuals examined, n > 30. **ii)**
1239 MIP with segmentation of the 23 *ago2-1*$^{+}$ cells. **Di)** *P53* and *UPPA* expression shows the
1240 two stem cell populations intermingled in 100% of individuals examined, n > 30 **ii)** MIP and
1241 segmentation reveal there are more *p53*$^{+}$ cells than *UPPA*$^{+}$ cells (in this larva,15 are *p53*$^{+}$ and

1242 9 are *UPPA*$^{+}$). Scale shown in C also applies to D.

1243

1244 **Figure 6. RNA velocity analysis of stem and tegument cells from the miracidium and**
1245 **sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage origin
1246 of cells and cell cluster identity from Seurat analysis. **B)** RNA velocity analysis flow field
1247 show the generalised direction of RNA velocity. **C)** Latent time analysis shows an estimated
1248 temporal-relationship between cells. **B)** and **C)** are based on the expression of both spliced
1249 and unspliced transcripts, and their expression dynamics across cells and genes. The phase
1250 plot, velocity, and expression were calculated for **D)** *p53-1*, **E)** *Zfp-1*, **F)** *p-53-2* and **G)** *eled*
1251 (no batch correction). Gene expression has been normalised (CPM) and log-transformed
1252 using scvelo(v. 0.2.4) For all of **D) - G)**, the phase plot (left plot) shows the proportion of
1253 spliced and unspliced transcripts in each cell, where each point is a cell and is coloured by the
1254 clusters in **B)**. The purple almond-shape overlaid represents the processes of transcription,
1255 splicing, and degradation, where this can be modelled. The dashed line shows the estimated
1256 steady state where RNA transcription is constant. The middle panel shows the RNA velocity,
1257 which for each gene is based on how the observations deviated from the estimated steady
1258 state towards induction or repression. The right panel shows gene expression. *p53-1* and *zfp-1*
1259 are predominantly expressed in the Delta/Phi-like miracidia stem cells and some miracidia
1260 tegument, and velocity indicates active expression of *p53-1* but downregulation of *zfp-1* in
1261 the stem cells. *p53-2* is most highly expressed in the Kappa-like miracidia stem cells, and
1262 velocity indicates this gene is being actively transcribed. *eled* expression is very low and only
1263 spliced transcripts have been detected, but is generally restricted to the Kappa-like miracidia
1264 stem cells.

1265 **Figure 7**. **A model for the fate of the two stem populations in the miracidium**. Adding
1266 single-cell data for miracidia stem cells (this study) to existing stem cell scenarios on
1267 *Schistosoma mansoni* development (Wang *et al*., 2018; Li *et al*., 2021) shows the continuum
1268 of the Kappa (κ) population from the miracidium, through the intra-molluscan stages to the
1269 juvenile and adult stages inside the mammalian host. Wang *et al*. (2018) proposed that the κ
1270 cells give rise to epsilon (*ε*), *eled*$^{+}$*,* cells in the juvenile primordial testes, ovaries, and
1271 vitellaria (germline), as well as in a gradient increasing toward the posterior growth zone
1272 (soma). They suggested that germ cells may be derived from ε-cells early in juvenile
1273 development, and *eled* is the earliest germline marker yet identified in schistosomes. This led
1274 to the idea that *S. mansoni* does not specify its germline until the juvenile stage (Wang *et al*.,

1275 2018) and a germline-specific regulatory program (including *eled, oc-1, akkn, nanos-1,*
1276 *boule*) was identified in intra-mammalian stages (Wang *et al*., 2018; Li *et al*., 2021). We
1277 show expression of these genes in κ stem cells in the miracidium. This suggests that after ~6
1278 days of embryogenesis, at hatching of the miracidium, the cells that may contribute to the
1279 germline might already be segregated into the κ population and the molecular regulatory
1280 program that differentiates somatic (delta/phi, $\delta/_{\varphi}$) and germ cell (κ) lineages is present.
1281 Furthermore, as *p53-2* plays a genotoxic stress response role in adult reproductive cells
1282 (Wendt *et al*., 2022), its expression in κ cells in the miracidia is another line of evidence that
1283 indicates that this population may contain the pluripotent stem cells that likely give rise to the
1284 germline.
1285

1286 **Figure 8**. **Tissue level segmentation of a miracidium reveals the contribution of each**
1287 **tissue to this simple larva**. *In situ* hybridisation using HCR (hybridisation chain reaction)
1288 enabled multiple tissue-level marker genes to be visualised simultaneously in the same larva.
1289 The nuclei of the cells of each tissue type were manually segmented using TrakEM2 in
1290 ImageJ (Cardona *et al*., 2012).
1291

1292 **Supplementary figure legends**
1293

1294 **Supplementary Figure 1. Contributions of sorted and unsorted cells to Seurat clusters.**
1295 This figure uses the same UMAP topology as presented in Figure 1C. **A)** Shows contribution
1296 of sorted cells to labelled clusters, **B)** shows contribution of unsorted cells to labelled
1297 clusters. **C)** UMAP where cells are coloured by sorted and unsorted treatment. **D)** UMAP
1298 where cells are coloured by sample of origin (samples 1 and 4 are sorted, samples 2 and 3 are
1299 unsorted).
1300

1301 **Supplementary Figure 2. Subclustering the Neuron 1 cluster using SAM revealed**
1302 **multiple distinct subpopulations**. **A)** The Neuron 1 cluster was analysed using Self-
1303 Assembling Manifold (SAM) algorithm, and Leiden clustering indicated 16 subpopulations.
1304 Library size was normalised and gene expression values were log-normalised using SAM
1305 (v1.0.1) and Scanpy (v1.8.2). **B)** The top five marker genes for each of the subpopulations.
1306 **C)** Expression of NPP PWamide, the top marker for subcluster 14. **D)** Expression of
1307 neuropeptide Y, the top marker for subcluster 13. **E)** Expression of neuropeptide F, the

1308 second top marker for subcluster 8. Although this cluster was difficult to define, this analysis
1309 suggests that there may be several small neural subpopulations within this cluster.
1310
1311 **Supplementary Figure 3. Neurons. A & B)** Top markers for cell cluster Neuron 1 include
1312 the neuropeptides precursor genes for Neuropeptide F, PWamide, and Neuropeptide Y. ISH
1313 of these genes shows expression in cells that make up the brain. **C)** G-alpha-q antibody
1314 staining reveals a girdle of 23 multi-ciliated nerve endings that sit between the 2nd and 3rd tier
1315 of ciliary plates. **D)** Neuron 4 marker Smp_319480 (putative mTORC1) transcripts are
1316 expressed at the base of these cilia; **E)** and some transcripts also reach the nuclei of circular
1317 body wall muscles. **F)** The expression of *p25 alpha* (which is a marker for ciliated cells in
1318 adult worms, Wendt *et al*., 2020) in the submuscular cells suggests that they may be ciliated.
1319

1320 **Supplementary Figure 4**. **Phylogeny of metazoan opsins.** There are four *Schistosoma*
1321 *mansoni* opsins that are expressed in two sub-clusters of Neuron 1; our analysis shows that
1322 three belong to the rhabdomeric opsin clade and the fourth is a peropsin. Scale bar shows
1323 branch lengths in expected amino acid substitutions per site. Shading of circles at internal
1324 nodes shows percentage support for the partition implied by that node from 1,000 non-
1325 parametric bootstrap samples.

1326

1327 **Supplementary Figure 5**. **The expression of two Polycystin genes and a Sox**
1328 **transcription factor in Neuron 4 cells. A)** Polycystin 1 and 2 are co-expressed in Neuron 4
1329 cells, orthologs in other animals form heterodimers in ciliated cells that function to detect
1330 flow sheer and vibrations (Kim *et al*., 2016; Calderon *et al*., 2018). Gene expression has been
1331 log-normalised and scaled using Seurat(v. 4.3.0). **B)** In planaria, a transcription factor Sox
1332 B1-2 regulates polycystin genes and together function in sensory neurons that detect flow.
1333 The co-expression of polycystin genes with a Sox transcription factor in Neuron 4 cells
1334 indicates that this transcription factor may regulate these polycystin genes in the miracidium.

1335

1336 **Supplementary Figure 6. STRING analysis of neural clusters and ciliary plates.**
1337 STRING analysis of the top 100 marker genes for each of the neural populations whose
1338 predicted networks have functional enrichment. **A)** Neuron 4, **B)** Neuron 5, **C)** Ciliary plates,

1339 **D)** Neuron 3. Neuron 1 is not a defined cluster and Neuron 2 showed no functional
1340 enrichment. Lines (edges) connecting nodes are based on evidence of the function of
1341 homologs. Functional enrichment (*FDR*) as provided by STRING. (PPI = predicted protein
1342 interaction). Minimum interaction (confidence) score of 0.7, corresponding to a high level of
1343 confidence, 0.4 medium level of confidence.
1344

1345 **Supplementary Figure 7. Gene expression by transcripts per million in the ciliary plates**
1346 **and gland cells.** Heatmap showing the gene expression, by transcripts per million, detected
1347 in the ciliary plates and gland cells hand-picked for plate-based scRNA-seq. Genes included;
1348 the top 20 significant genes between gland cell and ciliary plates, two neural markers, five
1349 VAL genes, five significant genes between gland cells and other cells, eight genes previously
1350 reported to produce secreted proteins in the miracidia (Wang *et al*., 2016; Wu *et al*., 2006),
1351 five top marker genes from the ciliary plate cluster.

1352

1353 **Supplementary Figure 8. STRING analysis of tegument and protonephridia clusters.**
1354 STRING analysis of the top 100 marker genes for **A)** tegument, **B)** protonephridia. Lines
1355 (edges) connecting nodes are based on evidence of the function of homologs. Functional
1356 enrichment (*FDR*) as provided by STRING. (PPI = predicted protein interaction). Minimum
1357 interaction (confidence) score of 0.7, corresponding to a high level of confidence, 0.4
1358 medium level of confidence.
1359

1360 **Supplementary Figure 9. STRING analysis of parenchyma clusters.** STRING analysis of
1361 the top 100 marker genes for **A)** Parenchyma 1, **B)** Parenchyma 2. Lines (edges) connecting
1362 nodes are based on evidence of the function of homologs. Functional enrichment (*FDR*) as
1363 provided by STRING (PPI = predicted protein interaction). Minimum interaction
1364 (confidence) score of 0.7, corresponding to a high level of confidence, 0.4 medium level of
1365 confidence.

1366

1367 **Supplementary Figure 10. Gene ontology (GO) analysis of marker genes identified in**
1368 **the miracidia cell clusters.** GO analysis was performed on each cluster using genes with a
1369 minimum AUC score of 0.7. Enriched biological process (BP) GO terms are shown for each
1370 cluster, and have been filtered to only show terms supported by a minimum of two genes.

1371 Top panels (left to right) show parenchyma, muscle, and tegument clusters, middle panels
1372 (left to right) show ciliary plate, neuron, and protonephridia clusters, and the bottom panel
1373 shows stem clusters. Full list of GO terms available in Supplementary Table 7.
1374
1375 **Supplementary Figure 11. Expression of known stem cell markers in miracidia cells.**
1376 **A-F)** Three stem cell classes were identified in the mother sporocyst stage and were named
1377 based upon their respective markers: Kappa (*klf*$^+$, *nanos-2*$^+$), Delta (*nanos-2*$^+$ and *fgfrA*$^+$,
1378 *fgfrB*$^+$ and *nanos-2*$^+$) and Phi (*fgfrA*$^+$and, *fgfrB*$^+$ and *hesl*$^+$), as well as pan-stem marker *ago2-*
1379 *1* (Wang *et al*., 2018; Safarti *et al*., 2021). In the miracidium, we identified seven stem cell
1380 clusters by expression of *ago2-1*: stem A and B are *klf*$^+$ and *nanos-2*$^+$ (i.e. Kappa-like), stem
1381 C and D are *nanos-2*$^+$, *fgfrA*$^+$, *fgfrB*$^+$, and *hesl*$^+$ (i.e. they resembled both Delta and Phi).
1382 Gene expression has been log-normalised and scaled using Seurat(v. 4.3.0).
1383
1384 **Supplementary Figure 12 - The effect of sex-linked genes on cell clustering.**
1385 *Schistosoma mansoni* has 7 pairs of autosomes and one pair of sex chromosomes; males have
1386 ZZ and females have ZW (Buddenborg *et al*., 2021). The sex chromosomes are composed of
1387 sex-specific regions that are unique to each chromosome flanked by pseudoautosomal regions
1388 that are common to both Z and W. The W-specific region (WSR) is a large, highly repetitive
1389 region that contains 38 genes, most of which are gametologs (genes with homologous copies
1390 on the WSR and ZSR). The Z-specific region (ZSR) contains 941 genes, of which only 35 are
1391 gametologs. **A & B)** The proportion of reads mapping to the WSR and ZSR across all cell
1392 type clusters shows **A)** putative male cells (Stem A and C i.e. those that lack WSR gene
1393 expression) and **B)** evidence of incomplete ZSR gene dosage compensation in the male stem
1394 cells. **C)** To determine whether any of the WSR genes were responsible for the clustering
1395 pattern, we excluded them from the mapping reference. The impact on the overall UMAP
1396 clustering was minimal. We repeated this excluding the **D)** ZSR genes, **E)** gametologues and
1397 **F)** ZSR genes, except the gametologues, in turn. In all four scenarios the stem cell clusters
1398 remained separated by sex. ZSR or WSR expression is therefore sufficient on their own to
1399 split these stem clusters in two.
1400
1401 **Supplementary Figure 13. Identification of new clusters after removing ZSR and WSR**
1402 **genes.** To identify whether stem cell clusters A and B, and C and D, still split down A/B and
1403 C/D lines despite the apparent collapsing of clusters after removing ZSR and WSR genes, we
1404 ran de novo clustering at several resolutions. Even once clusters were oversplitting, there was

1405 not a separation between stem cluster A/B and C/D. This was run using Seurat (v5.0.2), so
1406 the overall topology is not identical to that shown in Figure 5, though the same number of
1407 PCs are used. **A)** Shows the UMAP topology labelled by original cluster ID, **B)** shows the de
1408 novo clusters at resolution = 0.6, and **C)** shows the de novo clusters at resolution = 1.
1409
1410 **Supplementary Figure 14. Classification of cells based on percentage of reads mapping**
1411 **to W-specific region indicates incomplete dosage compensation of Z-specific region**
1412 **genes**. **A)** The percentage of reads mapping to W-specific genes by cluster. The blue line at
1413 0.2% indicates where the data were split into W+ and W- categories, i.e. Stem B and D
1414 contain cells with reads mapping to W-specific genes (and are likely cells from female
1415 miracidia), whereas most cells in Stem A and C do not have reads mapping to W-specific
1416 genes (cells from male miracidia). **B)** UMAP showing the clusters by W+ and W- categories.
1417 Almost complete separation can be seen in the main stem cell clusters, and the categories are
1418 mixed in most other clusters.
1419
1420 **Supplementary Figure 15. EdU labelling shows that there are no dividing cells during**
1421 **the free-swimming miracidia stage**. **A)** 4 hour EdU pulse on miracidia 0-6 hours post-
1422 hatching shows no EdU incorporation in any cells. **B)** Negative control. **C)** Positive control,
1423 three day EdU pulse on freshly-transformed mother sporocysts shows many dividing cells
1424 incorporating EdU, all treatments n = 30.
1425
1426 **Supplementary Figure 16. STRING analysis of the top 100 marker genes for the two**
1427 **stem cell populations in the miracidium (combining the female and male cells). A)** 61 of
1428 the top markers of Stem C and D combined (i.e. Delta/Phi-like) and **B)** 37 for Stem A and B
1429 combined (i.e. Kappa-like) form large predicted interactions enriched for translation,
1430 ribosome biogenesis and nucleic acid binding. Lines (edges) connecting nodes are based on
1431 evidence of the function of homologues. Functional enrichment (*FDR*) as provided by
1432 STRING. PPI = predicted protein interaction. Minimum interaction (confidence) score of 0.7,
1433 corresponding to a high level of confidence.
1434
1435 **Supplementary Figure 17. Differential gene expression between the two stem cell**
1436 **populations, Delta/Phi and Kappa**. **A)** Heatmap of top differentially expressed genes (by
1437 adjusted p-value) from each of Delta/Phi and Kappa populations show a high number of
1438 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

1473 as calculated for genes that are mentioned in the text: red outline = known molecular
1474 regulators of germline stem cells in adult *S. mansoni*. Blue outline = pluripotency gene. *Oc-1*
1475 expression is low with little velocity signal, but the expression is predominantly seen in the
1476 Kappa-like miracidia stem cells (and Stem G). *Irx* is also primarily expressed in the Kappa-
1477 like miracidia stem cells with velocity results indicating an accumulation of spliced
1478 transcripts. *Irx* transcripts are also detected in Stem G and the sporocyst stem/germinal cells.
1479 *Akkn* expression is low and largely seen in the Kappa-like and Stem G miracidia stem
1480 clusters. Lsm14-domain containing protein transcripts are most consistently expressed in the
1481 Kappa-like miracidia stem cells. *Klf* is most highly expressed in sporocyst stem/germinal
1482 cells, and the Kappa-like miracidia stem cluster. Gene expression has been normalised
1483 (CPM) and log-transformed using scvelo(v. 0.2.4).
1484
1485 **Supplementary Figure 21. RNA velocity analysis of all tissue types from the miracidium**
1486 **and sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage
1487 origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity
1488 analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis
1489 shows an estimated temporal-relationship between cells. Phase plot, velocity and expression
1490 as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst
1491 cells combined with miracidia sample 1 cells). *P53-1* is most highly expressed in the
1492 Delta/phi-like miracidia stem clusters and some of the miracidia tegument cells, and the
1493 velocity indicates active transcription. *Zfp-1* is expressed most highly in the same clusters,
1494 but velocity suggests that the majority of transcripts are spliced. *P53-2* expression is highest
1495 in the Kappa-like miracidia stem clusters, but the expression is detected in several other
1496 clusters, including muscle, neuron, and other stem clusters. Velocity indicates that this gene
1497 is being actively transcribed in the majority of these clusters. *Eled* expression is very low, but
1498 restricted to the Kappa-like miracidia stem clusters. Gene expression has been normalised
1499 (CPM) and log-transformed using scvelo(v. 0.2.4)
1500
1501 **Supplementary Figure 22. RNA velocity analysis of all tissue types from the miracidium**
1502 **and sporocyst shows the dynamics of described germline-related genes. A)** UMAP shows
1503 the life cycle stage origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)**
1504 RNA velocity analysis flow field shows the generalized direction of RNA velocity. Phase
1505 plot, velocity and expression as calculated for **D)** *oc-1*, **E)** *boule*, **F)** *irx,* **G)** *akkn,* and **H)** an
1506 FDF domain-containing protein (no batch correction, sporocyst cells combined with

1507 miracidia sample 1 cells.). *Oc-1* expression is low but generally specific to the two Kappa-
1508 like miracidia stem cell clusters, and Stem G. Only spliced *Boule* transcripts were detected,
1509 and with low expression, predominantly in the Stem B and Stem E clusters. *Irx* was most
1510 highly expressed in Kappa-like miracidia stem clusters, as well as Stem E, Stem G, and
1511 sporocyst stem/germinal cells. *Akkn* is also largely expressed in the Kappa-like miracidia
1512 stem clusters. The FDF domain-containing protein expression is low, but seen most in
1513 Kappa-like miracidia stem clusters, Stem G, and several neural clusters. Gene expression has
1514 been normalised (CPM) and log-transformed using scvelo(v. 0.2.4).
1515
1516
1517 **Supplementary Figure 23. RNAvelocity analysis of all tissue types from the miracidium**
1518 **and sporocyst show lineage-specific gene dynamics. A)** UMAP shows the life cycle stage
1519 origin of cells and **B)** shows cell cluster identity from Seurat analysis. **C)** RNA velocity
1520 analysis flow field shows the generalised direction of RNA velocity. **D)** Latent time analysis
1521 shows an estimated temporal-relationship between cells. Phase plot, velocity and expression
1522 as calculated for **E)** *p53-1*, **F)** *Zfp-1*, **G)** *p-53-2* and **H)** *eled* (no batch correction, sporocyst
1523 cells combined with miracidia sample 4 cells). *p53-1* and *zfp-1* are predominantly in the
1524 Delta/Phi-like miracidia stem cell clusters (Stem C and D) and also in the miracidia
1525 tegument. Velocity indicates that *p53-1* is being induced in the stem cells and repressed in the
1526 tegument, while *zfp-1* shows a gradient of velocity in the stem cells and repression in the
1527 tegument. *P53-2* is most highly expressed in the Kappa-like miracidia stem cells (Stem A and
1528 B) and Stem F, with velocity results indicating active transcription. *eled* expression was very
1529 low, but primarily in the Kappa-like miracidia stem cells (particularly Stem A). Gene
1530 expression has been normalised (CPM) and log-transformed using scvelo(v. 0.2.4).
1531
1532 **References**
1533

1534 1. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-
1535 cell RNA-seq artifacts: a molecular atlas of kidney development. Development.
1536 2017;144: 3625–3632. doi:10.1242/dev.151142

1537 2. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. In: R
1538 package version 2.54.0 [Internet]. 2023. doi:10.18129/B9.bioc.topGO

1539 3. Ashton PD, Harrop R, Shah B, Wilson RA. The schistosome egg: development and
1540 secretions. Parasitology. 2001;122: 329–338. doi:10.1017/s0031182001007351

1541 4. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to
1542 transient cell states through dynamical modeling. Nat Biotechnol. 2020;38:
1543 1408–1414. doi:10.1038/s41587-020-0591-3

1544 5. Bezares-Calderón LA, Berger J, Jasek S, Verasztó C, Mendes S, Gühmann M, et al.
1545 Neural circuitry of a polycystin-mediated hydrodynamic startle response for
1546 predator avoidance. Elife. 2018;7. doi:10.7554/eLife.36262

1547 6. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
1548 quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

1549 7. Buddenborg SK, Lu Z, Sankaranarayan G, Doyle SR, Berriman M. The stage- and
1550 sex-specific transcriptome of the human parasite Schistosoma mansoni. Sci
1551 Data. 2023;10: 775. doi:10.1038/s41597-023-02674-2

1552 8. Buddenborg SK, Tracey A, Berger DJ, Lu Z, Doyle SR, Fu B, et al. Assembled
1553 chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the
1554 evolution of its ZW sex-determination system. bioRxiv. 2021;
1555 2021.08.13.456314-2021.08.13.456314. doi:10.1101/2021.08.13.456314

1556 9. Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, et
1557 al. TrakEM2 software for neural circuit reconstruction. PLoS One. 2012;7:
1558 e38011–e38011. doi:10.1371/journal.pone.0038011

1559 10. Chernin E. Interference with the capacity of Schistosoma mansoni miracidia to
1560 infect the molluscan host. J Parasitol. 1968;54: 509–516.

1561 11. Choi HMT, Calvert CR, Husain N, Huss D, Barsi JC, Deverman BE, et al.
1562 Mapping a multiplexed zoo of mRNA expression. Development. 2016;143:
1563 3632–3637. doi:10.1242/dev.140137

1564 12. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J,
1565 et al. Third-generation in situ hybridization chain reaction: multiplexed,

1566 quantitative, sensitive, versatile, robust. Development. 2018;145.
1567 doi:10.1242/dev.165753

1568 13. Collins JJ 3rd, King RS, Cogswell A, Williams DL, Newmark PA. An atlas for
1569 Schistosoma mansoni organs and life-cycle stages using cell type-specific
1570 markers and confocal microscopy. PLoS Negl Trop Dis. 2011;5: e1009–e1009.
1571 doi:10.1371/journal.pntd.0001009

1572 14. Costain AH, MacDonald AS, Smits HH. Schistosome Egg Migration:
1573 Mechanisms, Pathogenesis and Host Immune Responses. Front Immunol.
1574 2018;9: 3042. doi:10.3389/fimmu.2018.03042

1575 15. Diaz Soria CL, Attenborough T, Lu Z, Fontenla S, Graham J, Hall C, Thompson S,
1576 Andrews TGR, Rawlinson KA, Berriman M, Rinaldi G. Single-cell
1577 transcriptomics of the human parasite Schistosoma mansoni first intra-
1578 molluscan stage reveals tentative tegumental and stem-cell regulators. Sci Rep.
1579 2024 Mar 12;14(1):5974. doi: 10.1038/s41598-024-55790-3. PMID: 38472267;
1580 PMCID: PMC10933418.

1581 16. Diaz Soria CL, Lee J, Chong T, Coghlan A, Tracey A, Young MD, et al. Single-
1582 cell atlas of the first intra-mammalian developmental stage of the human
1583 parasite Schistosoma mansoni. Nat Commun. 2020;11: 6411.
1584 doi:10.1038/s41467-020-20092-5

1585 17. Fishman L, Nechooshtan G, Erhard F, Regev A, Farrell JA, Rabani M. Single-cell
1586 temporal dynamics reveals the relative contributions of transcription and
1587 degradation to cell-type specific gene expression in zebrafish embryos. bioRxiv
1588 [Preprint]. 2023 Apr 21:2023.04.20.537620. doi: 10.1101/2023.04.20.537620.
1589 PMID: 37131717; PMCID: PMC10153228.

1590 18. Gąsiorowski L, Andrikou C, Janssen R, Bump P, Budd GE, Lowe CJ, et al.
1591 Molecular evidence for a single origin of ultrafiltration-based excretory organs.
1592 Curr Biol. 2021;31: 3629-3638.e2. doi:10.1016/j.cub.2021.05.057

1593 19. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al.
1594 Integrated analysis of multimodal single-cell data. Cell. 2021;184: 3573-
1595 3587.e29. doi:10.1016/j.cell.2021.04.048

1596 20. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016:
1597 expanding to enable helminth genomic research. Nucleic Acids Res. 2016;44:
1598 D774-80. doi:10.1093/nar/gkv1217

1599 21. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a
1600 comprehensive resource for helminth genomics. Mol Biochem Parasitol.
1601 2017;215: 2–10. doi:10.1016/j.molbiopara.2016.11.005

1602 22. IHME. 2019 Global Burden of Disease (GBD) study [Internet]. 2019. Available:
1603 https://vizhub.healthdata.org/gbd-results?params=gbd-api-2019-
1604 permalink/d30124bc2eaad3d977583362d4d37591 [vizhub.healthdata.org]

1605 23. Jurberg AD, Gonçalves T, Costa TA, de Mattos ACA, Pascarelli BM, de Manso
1606 PPA, et al. The embryonic development of Schistosoma mansoni eggs: proposal
1607 for a new staging system. Dev Genes Evol. 2009;219: 219–234.
1608 doi:10.1007/s00427-009-0285-9

1609 24. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang T-H, Liu Y, et al. Notum
1610 deacylates Wnt proteins to suppress signalling activity. Nature. 2015;519: 187–
1611 192. doi:10.1038/nature14259

1612 25. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell
1613 differential expression analysis. Nat Methods. 2014;11: 740–742.
1614 doi:10.1038/nmeth.2967

1615 26. Kim S, Nie H, Nesin V, Tran U, Outeda P, Bai C-X, et al. The polycystin complex
1616 mediates Wnt/Ca2+ signalling. Nat Cell Biol. 2016;18: 752–764.
1617 doi:10.1038/ncb3363

1618 27. Kumar P, Tan Y, Cahan P. Understanding development and stem cells using single
1619 cell-based analyses of gene expression. Development. 2017;144: 17–32.
1620 doi:10.1242/dev.133058

1621 28. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al.
1622 RNA velocity of single cells. Nature. 2018;560: 494–498. doi:10.1038/s41586-
1623 018-0414-6

1624 29. Li P, Nanes Sarfati D, Xue Y, Yu X, Tarashansky AJ, Quake SR, et al. Single-cell
1625 analysis of Schistosoma mansoni identifies a conserved genetic program
1626 controlling germline stem cell fate. Nat Commun. 2021;12: 485.
1627 doi:10.1038/s41467-020-20794-w

1628 30. Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba
1629 A. Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem
1630 Cells that Become Activated upon Brain Injury. Cell Stem Cell. 2015;17: 329–
1631 340. doi:10.1016/j.stem.2015.07.002

1632 31. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC, et al.
1633 EmptyDrops: distinguishing cells from empty droplets in droplet-based single-
1634 cell RNA sequencing data. Genome Biol. 2019;20: 63. doi:10.1186/s13059-
1635 019-1662-y

1636 32. Mair GR, Maule AG, Day TA, Halton DW. A confocal microscopical study of the
1637 musculature of adult Schistosoma mansoni. Parasitology. 2000;121 (Pt 2: 163–
1638 170. doi:10.1017/s0031182099006174

1639 33. Maldonado JF, Acosta-Matienzo J. Biological Studies on the Miracidium of
1640 Schistosoma Mansoni. Am J Trop Med. Arlington VA, USA: The American
1641 Society of Tropical Medicine and Hygiene; 1948;s1-28: 645–657.
1642 doi:10.4269/ajtmh.1948.s1-28.645

1643 34. Mann VH, Morales ME, Rinaldi G, Brindley PJ. Culture for genetic manipulation
1644 of developmental stages of Schistosoma mansoni. Parasitology. 2010;137: 451–
1645 462. doi:10.1017/S0031182009991211

1646 35. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in
1647 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell
1648 Syst. 2019;8: 329-337.e4. doi:https://doi.org/10.1016/j.cels.2019.03.003
1649 [doi.org]

1650 36. McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou X-N.
1651 Schistosomiasis. Nat Rev Dis Prim. 2018;4: 13. doi:10.1038/s41572-018-0013-
1652 8

1653 37. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KH (Joseph), et al. Modular,
1654 efficient and constant-memory single-cell RNA-seq preprocessing. Nat
1655 Biotechnol. 2021;39: 813–818. doi:10.1038/s41587-021-00870-2

1656 38. Moses L LP. BUSpaRse: kallisto | bustools R utilities [Internet]. 2023.
1657 doi:10.18129/B9.bioc.BUSpaRse

1658 39. Nanes Sarfati D, Li P, Tarashansky AJ, Wang B. Single-cell deconstruction of
1659 stem-cell-driven schistosome development. Trends Parasitol. 2021;37: 790–802.
1660 doi:https://doi.org/10.1016/j.pt.2021.03.005 [doi.org]

1661 40. Orosz F, Ovádi J. TPPP orthologs are ciliary proteins. FEBS Lett. 2008;582: 3757–
1662 3764. doi:10.1016/j.febslet.2008.10.011

1663 41. Pan SC. The fine structure of the miracidium of Schistosoma mansoni. J Invertebr
1664 Pathol. 1980;36: 307–372. doi:https://doi.org/10.1016/0022-2011(80)90040-3
1665 [doi.org]

1666 42. Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body
1667 axis. Cell. 2009;139: 1056–1068. doi:10.1016/j.cell.2009.11.035

1668 43. Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, et al. Dosage
1669 Compensation throughout the Schistosoma mansoni Lifecycle: Specific
1670 Chromatin Landscape of the Z Chromosome. Genome Biol Evol. 2019;11:
1671 1909–1922. doi:10.1093/gbe/evz133

1672 44. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of
1673 RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14: 687–
1674 690. doi:10.1038/nmeth.4324

1675 45. Poteaux P, Gourbal B, Duval D. Time series analysis of tegument ultrastructure of
1676 in vitro transformed miracidium to mother sporocyst of the human parasite

1677 Schistosoma mansoni. Acta Trop. 2023 Apr;240:106840. doi:
1678 10.1016/j.actatropica.2023.106840. Epub 2023 Jan 18. PMID: 36681315.

1679 46. Scimone ML, Cote LE, Reddien PW. Orthogonal muscle fibres have different
1680 instructive roles in planarian regeneration. Nature. 2017;551: 623–628.
1681 doi:10.1038/nature24660

1682 47. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-Cell RNA-
1683 Seq with Waterfall Reveals Molecular Cascades underlying Adult
1684 Neurogenesis. Cell Stem Cell. 2015;17: 360–372.
1685 doi:https://doi.org/10.1016/j.stem.2015.07.013 [doi.org]

1686 48. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, et al.
1687 Stem cell differentiation trajectories in Hydra resolved at single-cell resolution.
1688 Science (80-). American Association for the Advancement of Science;
1689 2019;365: eaav9314–eaav9314. doi:10.1126/science.aav9314

1690 49. Steger J, Cole AG, Denner A, Lebedeva T, Genikhovich G, Ries A, Reischl R,
1691 Taudes E, Lassnig M, Technau U. Single-cell transcriptomics identifies
1692 conserved regulators of neuroglandular lineages,Cell Reports, 40, 12,2022,
1693 111370, https://doi.org/10.1016/j.celrep.2022.111370.

1694 50. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al.
1695 STRING v11: protein–protein association networks with increased coverage,
1696 supporting functional discovery in genome-wide experimental datasets. Nucleic
1697 Acids Res. 2019;47: D607–D613. doi:10.1093/nar/gky1131

1698 51. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al.
1699 Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined
1700 Factors. Cell. 2007;131: 861–872. doi:https://doi.org/10.1016/j.cell.2007.11.019
1701 [doi.org]

1702 52. Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, et al.
1703 Mapping single-cell atlases throughout Metazoa unravels cell type evolution.
1704 Shalek AK, Barkai N, editors. Elife. eLife Sciences Publications, Ltd; 2021;10:
1705 e66747–e66747. doi:10.7554/eLife.66747

1706 53. Tarashansky AJ, Xue Y, Li P, Quake SR, Wang B. Self-assembling manifolds in
1707 single-cell RNA sequencing data. Shalek AK, Barkai N, Yanai I, editors. Elife.
1708 eLife Sciences Publications, Ltd; 2019;8: e48994–e48994.
1709 doi:10.7554/eLife.48994

1710 54. Valsecchi CIK, Basilicata MF, Semplicio G, Georgiev P, Gutierrez NM, Akhtar A.
1711 Facultative dosage compensation of developmental genes on autosomes in
1712 Drosophila and mouse embryonic stem cells. Nat Commun. 2018;9: 3626.
1713 doi:10.1038/s41467-018-05642-2

1714 55. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al.
1715 Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2023;
1716 doi:10.1038/s41587-023-01773-0

1717 56. Verasztó C, Jasek S, Gühmann M, Shahidi R, Ueda N, Beard JD, et al. Whole-
1718 animal connectome and cell-type complement of the three-segmented
1719 Platynereis dumerilii larva. bioRxiv. 2020; 2020.08.21.260984-
1720 2020.08.21.260984. doi:10.1101/2020.08.21.260984

1721 57. Vergara HM, Pape C, Meechan KI, Zinchenko V, Genoud C, Wanner AA, et al.
1722 Whole-body integration of gene expression and single-cell morphology. Cell.
1723 2021;184: 4819-4837.e22. doi:https://doi.org/10.1016/j.cell.2021.07.017
1724 [doi.org]

1725 58. Vogg MC, Owlarn S, Pérez Rico YA, Xie J, Suzuki Y, Gentile L, et al. Stem cell-
1726 dependent formation of a functional anterior regeneration pole in planarians
1727 requires Zic and Forkhead transcription factors. Dev Biol. 2014;390: 136–148.
1728 doi:https://doi.org/10.1016/j.ydbio.2014.03.016 [doi.org]

1729 59. Wang B, Collins III JJ, Newmark PA. Functional genomic characterization of
1730 neoblast-like stem cells in larval Schistosoma mansoni. Sánchez Alvarado A,
1731 editor. Elife. eLife Sciences Publications, Ltd; 2013;2: e00768–e00768.
1732 doi:10.7554/eLife.00768

1733 60. Wang B, Lee J, Li P, Saberi A, Yang H, Liu C, et al. Stem cell heterogeneity
1734 drives the parasitic life cycle of Schistosoma mansoni. Sánchez Alvarado A,

1735 editor. Elife. eLife Sciences Publications, Ltd; 2018;7: e35449–e35449.
1736 doi:10.7554/eLife.35449

1737 61. Wang T, Zhao M, Rotgans BA, Strong A, Liang D, Ni G, et al. Proteomic Analysis
1738 of the Schistosoma mansoni Miracidium. PLoS One. Public Library of Science;
1739 2016;11: e0147247–e0147247.

1740 62. Wendt GR, Collins JNR, Pei J, Pearson MS, Bennett HM, Loukas A, et al.
1741 Flatworm-specific transcriptional regulators promote the specification of
1742 tegumental progenitors in Schistosoma mansoni. Patel NH, editor. Elife. eLife
1743 Sciences Publications, Ltd; 2018;7: e33221–e33221. doi:10.7554/eLife.33221

1744 63. Wendt GR, Shiroor DA, Adler CE, Collins JJ. Convergent evolution of a genotoxic
1745 stress response in a parasite-specific p53 homolog. Proc Natl Acad Sci.
1746 Proceedings of the National Academy of Sciences; 2022;119: e2205201119–
1747 e2205201119. doi:10.1073/pnas.2205201119

1748 64. Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, et al. A single-
1749 cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood
1750 feeding. Science (80-). 2020;369. doi:10.1126/science.abb7709

1751 65. WHO. Schistosomiasis [Internet]. 2023. Available: https://www.who.int/news-
1752 room/fact-sheets/detail/schistosomiasis [who.int]

1753 66. Witchley JN, Mayer M, Wagner DE, Owen JH, Reddien PW. Muscle Cells
1754 Provide Instructions for Planarian Regeneration. Cell Rep. 2013;4: 633–641.
1755 doi:https://doi.org/10.1016/j.celrep.2013.07.022 [doi.org]

1756 67. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression
1757 data analysis. Genome Biol. 2018;19: 15. doi:10.1186/s13059-017-1382-0

1758 68. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al.
1759 Massively parallel digital transcriptional profiling of single cells. Nat Commun.
1760 2017;8: 14049. doi:10.1038/ncomms14049

1761

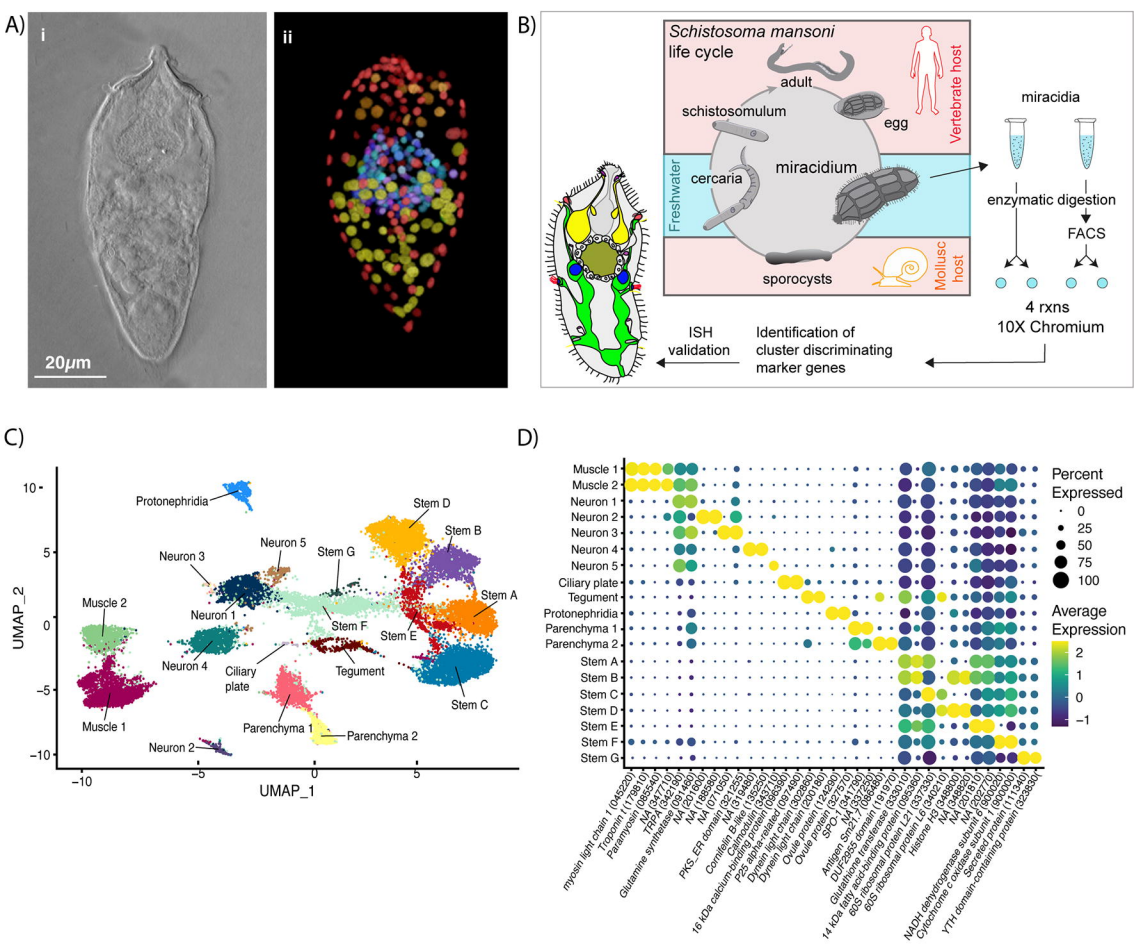
1762

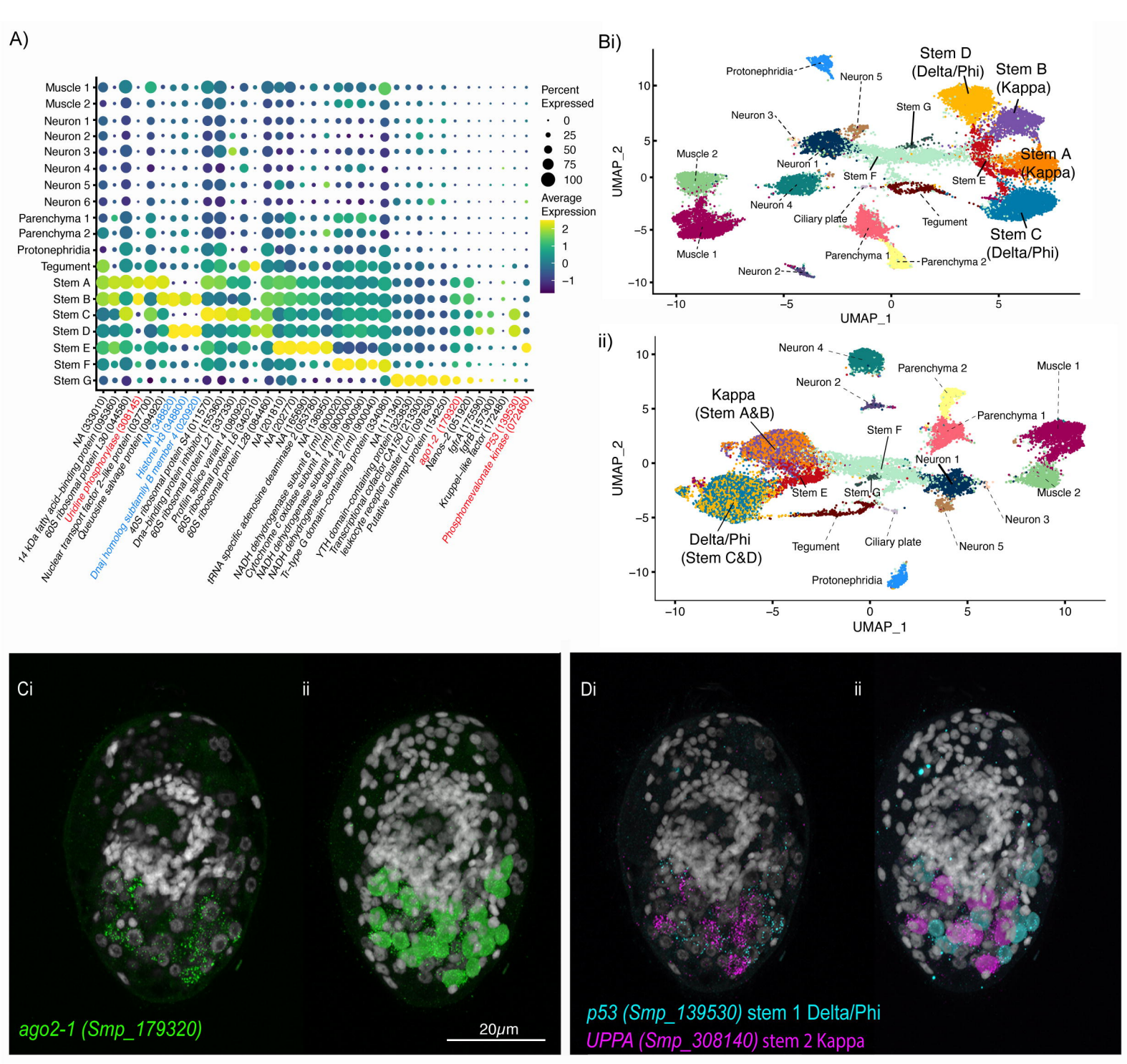
1763
1764
1765 **Supplementary Video legends**
1766
1767 **Supplementary Video 1** – Confocal z-stack of wholemount *in situ* hybridisation (WISH) of
1768 pan-muscle marker *paramyosin* (*PRM* Smp_085540) (cyan), counterstained with phalloidin
1769 (green) and DAPI (white), reveals which *PRM*$^{+}$ nuclei belong to the circular and longitudinal
1770 body wall muscles.
1771
1772 **Supplementary Video 2 –** Segmentation of *paramyosin* (Smp_085540) (cyan) positive cells
1773 (using the z-stack from Supplementary Figure 2); the nuclei of the circular body wall muscles
1774 are highlighted in orange, the nuclei of the longitudinal body wall muscles are highlighted in
1775 yellow.
1776
1777 **Supplementary Video 3 -** Confocal z-stack of WISH of muscle cluster 1 markers *Kunitz-*
1778 *type protease inhibitor* (Smp_052230) (cyan) and *Wnt-11-1* (Smp_156540) (magenta),
1779 counterstained with phalloidin (green) and DAPI (white), show expression in circular body
1780 wall muscles.
1781
1782 **Supplementary Video 4 –** Maximum intensity projection from confocal z-stack of WISH of
1783 muscle cluster 1 markers *notum* (Smp_015820) (cyan) and *Wnt-11-1* (Smp_156540)
1784 (magenta) counterstained with phalloidin (green) showing expression at opposite poles.
1785
1786 **Supplementary Video 5 -** Confocal z-stack of WISH of muscle cluster 2 markers
1787 *calcium activated potassium channel* (Smp_156150) (cyan) and *myoD* (Smp_167400)
1788 (magenta), counterstained with phalloidin (green) and DAPI (white), shows expression in
1789 longitudinal body wall muscles.
1790
1791 **Supplementary Video 6 -** Confocal z-stack of WISH of markers for neuron cluster 2
1792 (Smp_201600) (magenta) and neuron 3 (Smp_071050)(cyan), counterstained with DAPI
1793 (white), shows expression of both genes in cells whose nuclei form part of the nuclear rind of
1794 the brain.
1795

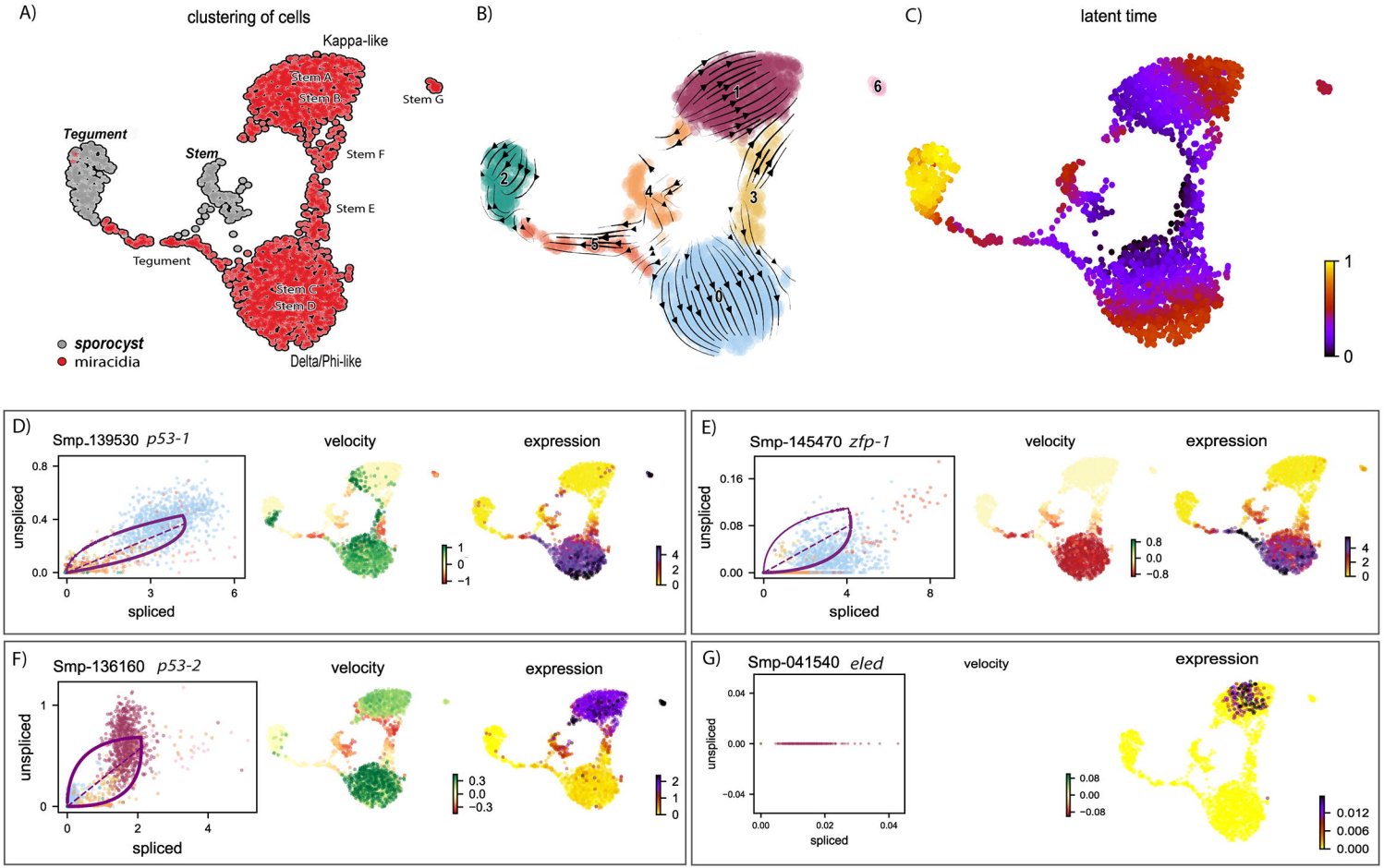
1796 **Supplementary Video 7 -** Confocal z-stack of WISH of ciliary plate marker *16 kDa calcium*
1797 *binding protein* (Smp_096390), counterstained with phalloidin (green) and DAPI (white),
1798 shows expression in the ciliary plates and six previously undescribed cells that we term sub-
1799 muscular cells. The nuclei of the sub-muscular cells sit beneath the body wall muscles, and
1800 each cell has a protrusion (possible cilia) that extends beyond the body wall muscles, between
1801 the first and second tiers of ciliary plates, and out to the exterior of the larva.
1802
1803 **Supplementary Video 8 –** Maximum intensity projection of confocal z-stack of WISH of
1804 ciliary plate marker *16 kDa calcium binding protein* (Smp_096390), counterstained with
1805 phalloidin (green) and DAPI (white), shows expression in ciliary plates and sub-muscular
1806 cells.
1807
1808 **Supplementary Video 9 –** Maximum intensity projection from confocal z-stack of WISH of
1809 neuron 2 marker Smp_319480 (magenta), counterstained with DAPI (white), showing
1810 expression in clusters of cells posterior to the brain that extend into the brain and out to the
1811 periphery.
1812
1813 **Supplementary Video 10 –** Maximum intensity projection from confocal z-stack of WISH
1814 of neuron 4 marker Smp_319480 (magenta) and ciliary plate marker *16 kDa calcium binding*
1815 *protein* (Smp_096390) (green), counterstained with DAPI (white), showing that the clusters
1816 of neuron 4 cells reach the periphery at the latitude between the second and third tiers of
1817 ciliary plates.
1818
1819 **Supplementary Video 11 -** Maximum intensity projection from confocal z-stack of WISH of
1820 neuron 5 marker Smp_319480 (cyan), counterstained with DAPI (white), showing expression
1821 in cells anterior to the brain and in and around the paired lateral glands and their secretory
1822 ducts.
1823
1824 **Supplementary Video 12 -** Maximum intensity projection from confocal z-stack of WISH of
1825 protonephridia marker *ShKT-domain containing protein* (Smp_335600) (magenta),
1826 counterstained with DAPI (white). This shows the transcripts in a bilaterally symmetrical
1827 pattern distributed along the s-shaped path of an excretory tubule that connects the anterior
1828 and posterior flame cells.
1829

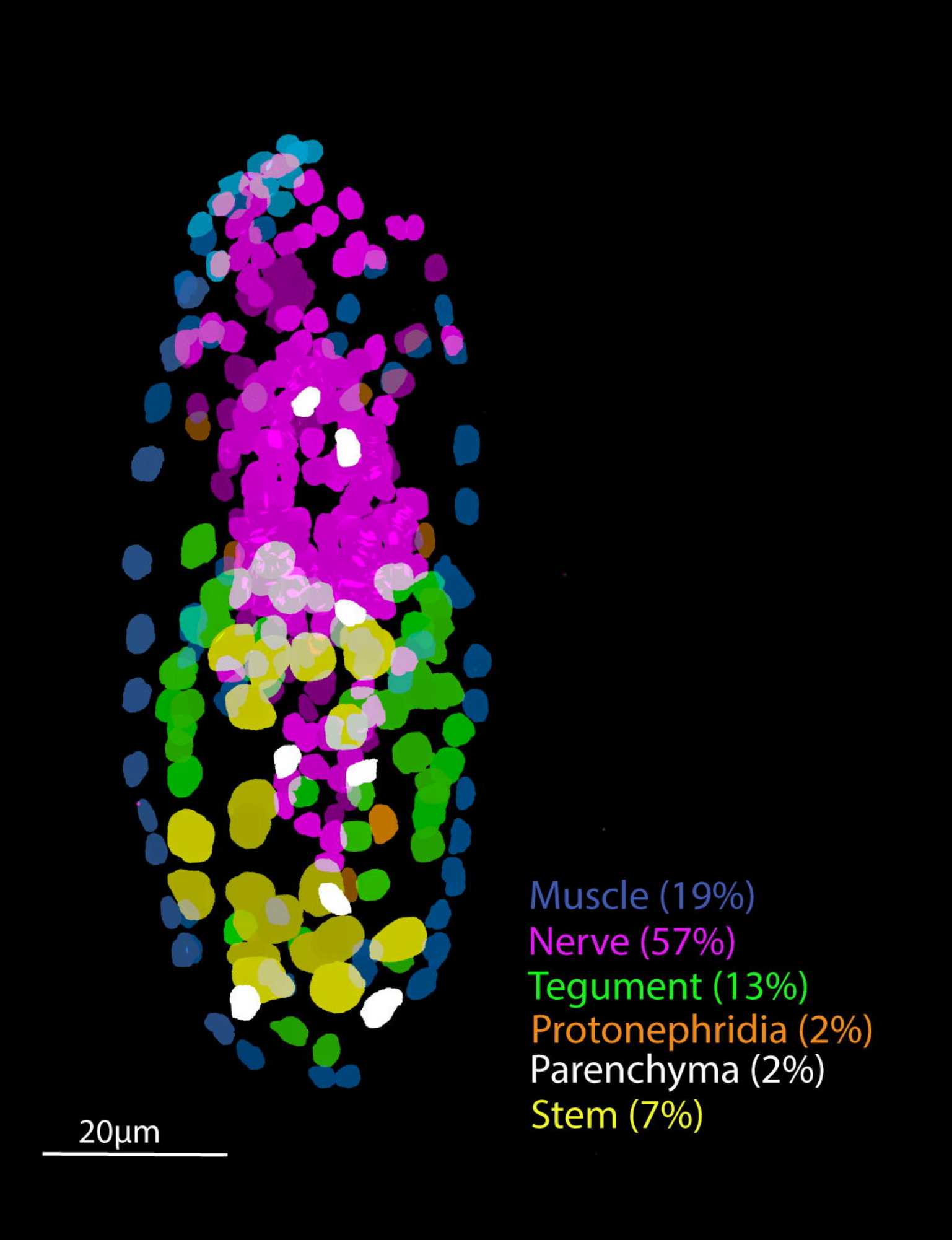
1830 **Supplementary Video 13 -** Confocal z-stack of WISH of markers for parenchyma 1 and 2
1831 (Smp_318890) (green) and stem cells *ago2-1* (Smp_179320) (magenta), counterstained with
1832 DAPI (white), shows the long cytoplasmic protrusions of the parenchymal cells reaching
1833 between the stem and other cells.
1834
1835 **Supplementary Video 14 -** Maximum intensity projection from confocal z-stack of WISH of
1836 marker for parenchyma 1 and 2 (Smp_318890) (green) counterstained with DAPI (white),
1837 shows the distribution of the parenchymal cells and their reach in the intercellular space of
1838 the larva.
1839
1840 **Supplementary Video 15 -** Confocal z-stack of WISH of markers for tegument *Meg6*
1841 (Smp_163710) (magenta) and *dynein light chain* (Smp_200180) (cyan), counterstained with
1842 phalloidin (green) and DAPI (white), shows the nuclei sit below the body wall muscle, and
1843 cytoplasmic protrusions reach between muscle filaments and form the epidermal ridges
1844 between the ciliary plates.
1845
1846 **Supplementary Video 16 -** Confocal z-stack of WISH of markers for stem clusters Delta/Phi
1847 *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (green) and stem E
1848 *Phosphomevalonate kinase* (Smp_072460) (magenta), counterstained with DAPI (white),
1849 shows there are no cells that are *pmk*-positive but negative for *UPPA* and *p53*. Stem E cells
1850 cannot be validated *in situ*.
1851
1852 **Supplementary Video 17 -** Confocal z-stack of WISH of markers for stem clusters Delta/Phi
1853 *p53* (Smp_139530) (cyan), Kappa *UPPA* (Smp_308145) (magenta) and the pan-stem marker
1854 *ago2-1* (Smp_179320) (green), counterstained with DAPI (white), shows there are no cells
1855 that are *ago2-1* positive but negative for *UPPA* and *p53* that could be Stem F and G cells.
1856
1857 **Supplementary Video 18 -** Confocal z-stack of multiplexed WISH of tissue markers for
1858 stem *ago2-1* (Smp_179320) (green), muscle *paramyosin* (Smp_085540) (cyan) and nerve
1859 *complexin* (Smp_050220)(magenta).
1860
1861

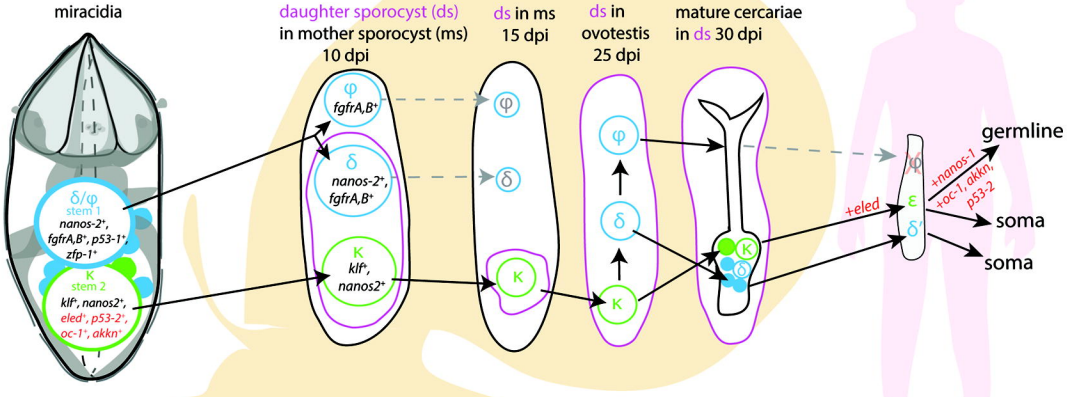
1862 **Supplementary Video 19 -** Confocal z-stack of multiplexed WISH of tissue markers for
1863 parenchyma (Smp_318890) (green), stem *ago2-1* (Smp_179320) (cyan) and tegument *Meg6*
1864 (Smp_163710) (magenta).
1865
1866 **Supplementary Video 20 -** Confocal z-stack of multiplexed WISH of tissue markers for
1867 tegument *Meg6* (Smp_163710) (green) and protonephridia *ShKT-domain containing protein*
1868 (Smp_335600) (magenta).
1869
1870

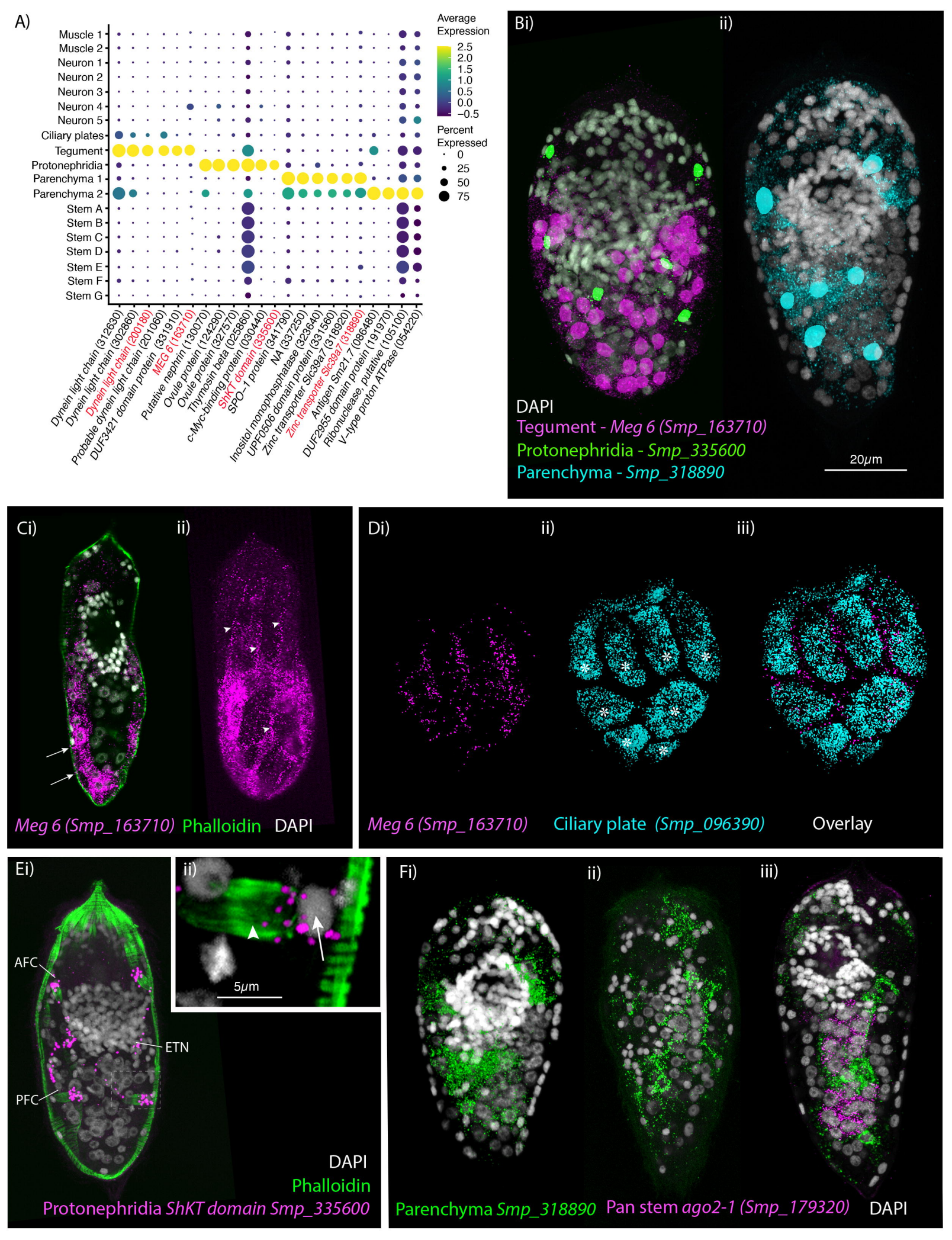

A)
i
ii
20μm
B)
Schistosoma mansoni
life cycle
adult
schistosomulum
egg
Vertebrate host
Freshwater
cercaria
miracidium
sporocysts
Mollusc host
miracidia
enzymatic digestion
FACS
4 rxns
10X Chromium
ISH
validation
Identification of
cluster discriminating
marker genes
C)
Protonephridia
Stem D
Stem B
Neuron 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 (17981)
Paramyosin (085540)
NA (342710)
TRPA (342180)
Glutamine synthetase (091480)
NA (201600)
NA (188580)
NA (071050)
PKS_ER domain (321255)
NA (319480)
Cornifelin B-like (135250)
Calmodulin (349310)
16 kDa calcium-binding protein (085390)
P25 alpha-related (097490)
Dynein light chain (302860)
Dynein light chain (200180)
Ovule protein (124290)
Ovule protein (322270)
SPO-1 (341790)
NA (335250)
Antigen Sm21.7 (086480)
DUF2955 domain (191970)
Glutathione transferase (333010)
14 kDa fatty acid-binding protein (305360)
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 (900030)
Secreted protein (113340)
YTH domain-containing protein (323830)

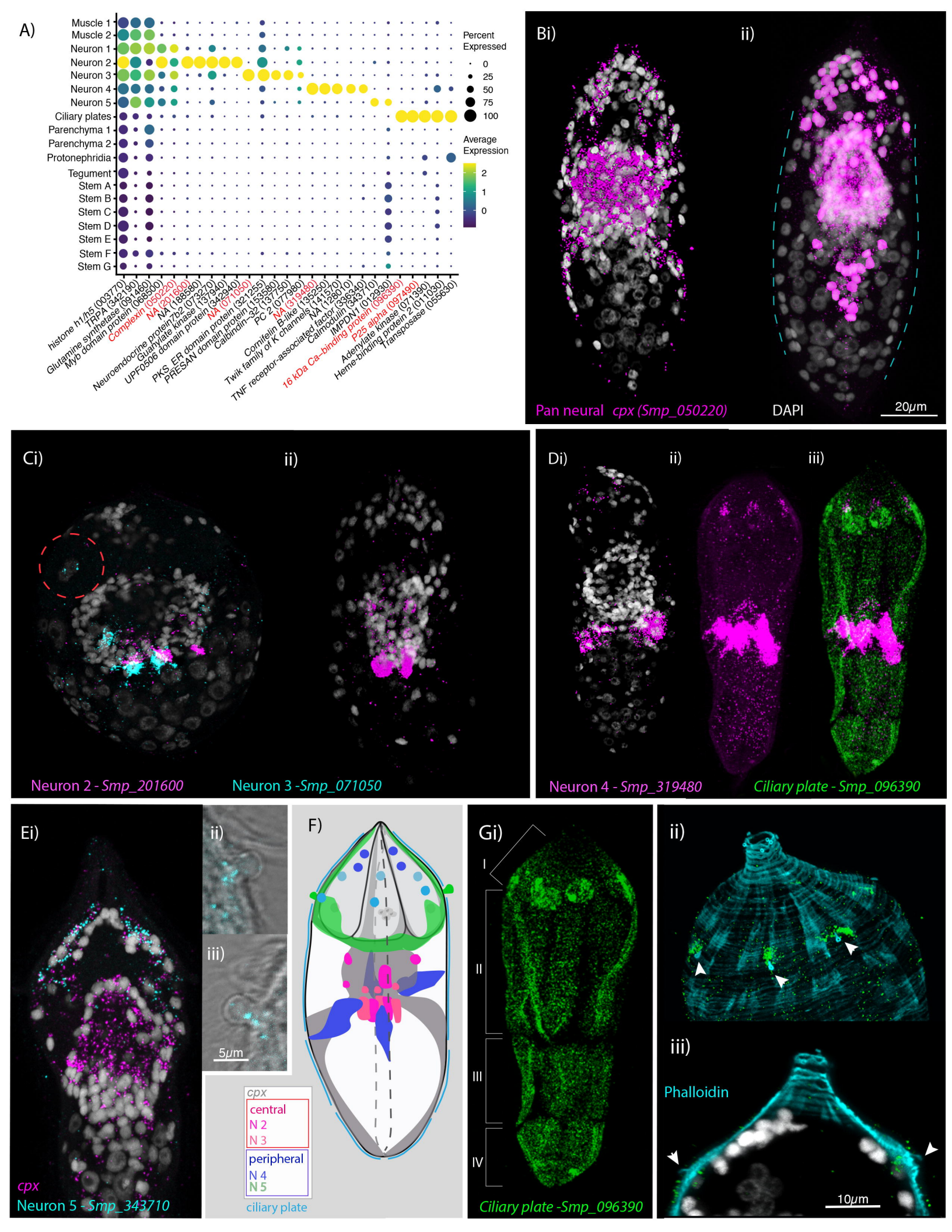

A)
Muscle 1
Muscle 2
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Neuron 6
Parenchyma 1
Parenchyma 2
Protonephridia
Tegument
Stem A
Stem B
Stem C
Stem D
Stem E
Stem F
Stem G
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
−1
NA (339010)
14 kDa fatty acid–binding protein (095360)
60S ribosomal 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

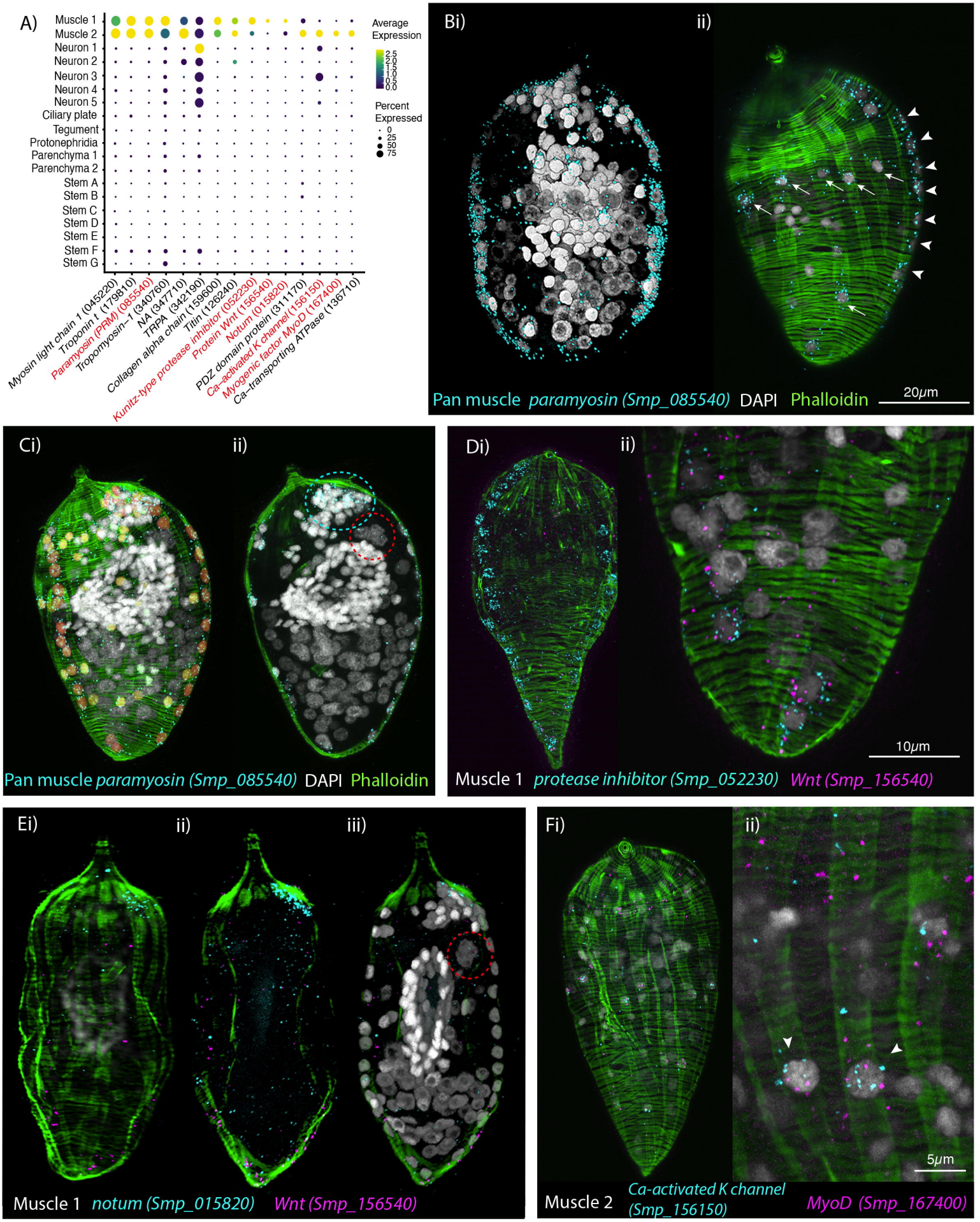

A)
clustering of cells
Kappa-like
Stem A
Stem B
Stem G
Tegument
Stem
Stem F
Stem E
Tegument
Stem C
Stem D
sporocyst
miracidia
Delta/Phi-like
B)
1
2
3
4
5
0
6
C)
latent time
1
0
D)
Smp_139530 p53-1
velocity
expression
unspliced
spliced
E)
Smp-145470 zfp-1
velocity
expression
unspliced
spliced
F)
Smp-136160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


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


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


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
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 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)