



OPEN

Single-cell transcriptomics of the human parasite *Schistosoma mansoni* first intra-molluscan stage reveals tentative tegumental and stem-cell regulators

Carmen L. Diaz Soria^{1,7}, Teresa Attenborough^{1,2,7}, Zhigang Lu¹, Santiago Fontenla³, Jennie Graham¹, Christopher Hall¹, Sam Thompson¹, Toby G. R. Andrews⁴, Kate A. Rawlinson^{1,5}, Matthew Berriman^{1,2}✉ & Gabriel Rinaldi^{1,6}✉

Schistosomiasis is a major Neglected Tropical Disease, caused by the infection with blood flukes in the genus *Schistosoma*. To complete the life cycle, the parasite undergoes asexual and sexual reproduction within an intermediate snail host and a definitive mammalian host, respectively. The intra-molluscan phase provides a critical amplification step that ensures a successful transmission. However, the cellular and molecular mechanisms underlying the development of the intra-molluscan stages remain poorly understood. Here, single cell suspensions from *S. mansoni* mother sporocysts were produced and sequenced using the droplet-based 10X Genomics Chromium platform. Six cell clusters comprising two tegument, muscle, neuron, parenchyma and stem/germinal cell clusters were identified and validated by in situ hybridisation. Gene Ontology term analysis predicted key biological processes for each of the clusters, including three stem/germinal sub-clusters. Furthermore, putative transcription factors predicted for stem/germinal and tegument clusters may play key roles during parasite development and interaction with the intermediate host.

Schistosomiasis, infection with parasitic flatworms in the genus *Schistosoma*, remains a major neglected tropical disease with more than 240 million people affected worldwide and more than 700 million at risk of infection in endemic areas^{1,2}. To date, a single drug (Praziquantel) is in use; however, this drug is effective only against adult worms, it does not prevent reinfection and drug resistance may be emerging in the field³. One approach that may lead to novel control strategies is to gain a better understanding of the mechanisms underlying life cycle progression, including the cells and their transcriptomic signatures across developmental stages^{4,5}.

Schistosoma mansoni eggs laid by adult worm pairs dwelling in the portal system of the mammalian host traverse the intestinal wall and pass with the faeces into the environment. In contact with fresh water, from the eggs hatch free-swimming larvae (miracidia) that seek, infect a suitable snail, and transform into mother sporocysts, the first intra-molluscan developmental stage. Within the mother sporocyst, groups of stem cells (historically termed 'germinal cells') start to proliferate and differentiate to develop into daughter sporocysts⁶. By 5 days post infection (5 dpi), developing daughter sporocysts, initially spherical in shape, start to grow within the brood chamber, elongate and become surrounded by a primitive epithelium derived from the mother sporocyst tegument⁶. By ~15 dpi they have acquired the definitive vermiform shape containing germinal cells closely packed in the medial part of the body⁶. The mature daughter sporocysts escape from the mother sporocyst, migrate through the snail tissue to the digestive gland area and start to produce whole cercariae from single germinal cells following a second round of embryogenesis⁷. Therefore, from a single miracidium, hundreds to thousands

¹Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK. ²School of Infection and Immunity, College of Medical, Veterinary and Life Sciences, University of Glasgow, 120 University Place, Glasgow G12 8TA, UK. ³Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Montevideo, Uruguay. ⁴The Francis Crick Institute, London, UK. ⁵Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, USA. ⁶Department of Life Sciences, Aberystwyth University, Edward Llwyd Building, Penglais Campus, Aberystwyth SY23 3DA, UK. ⁷These authors contributed equally: Carmen L. Diaz Soria and Teresa Attenborough. ✉email: Matt.Berriman@glasgow.ac.uk; gabriel.rinaldi@aber.ac.uk

of clonal human-infective cercariae are produced. Extensive knowledge gathered for several decades through detailed histological, and electron microscopy-based studies paved the way towards an understanding of the parasite progression within the snail^{8–11}. This knowledge can now be scrutinised using current molecular and ‘omics’ technologies¹². Shining new light on the cellular and molecular basis of this parasite expansion strategy is critical not only to discover novel aspects of trematode developmental biology, but also to reveal targets for control⁴.

Single-cell transcriptome sequencing (scRNA-seq) has been employed to define cellular subtypes by revealing their specific transcriptional signatures. Compared with so-called ‘bulk RNA-seq’ studies of whole organisms or tissues, scRNA-seq has exceptional resolving power, being able to detect genes expressed in just a few cells or with low expression levels, but also reveals the stochastic nature of gene expression in individual cells¹³. Single cell transcriptomics have been used in several systems to understand diverse biological processes, such as cell differentiation, tissue specification and development as well as to generate cell “atlases” based on the scRNA-seq profiles across different tissues¹⁴. Studies have employed scRNA-seq in *S. mansoni*^{4,5} to generate cell atlases for male and female adult worms¹⁵, mixed-sex schistosomula¹⁶, the first intra-mammalian developmental stage, and more recently mixed-sex miracidia¹⁷. In addition, scRNA-seq in schistosomes has been employed to define and functionally characterise stem cell populations driving the development of both intra-molluscan¹⁸ and intra-mammalian stages¹⁹.

Important contributions to our current understanding of the developmental biology of schistosome intra-molluscan stages have been made using ‘bulk RNA-seq’²⁰. In particular, two germinal cell lineages with distinct proliferation properties had previously been identified and functionally characterised by ‘bulk transcriptomics’ and RNAi, respectively²⁰. However, there is a critical lack of transcriptomic data and knowledge of gene regulatory networks at the single cell level. Wang and collaborators pioneered work in this area by sequencing 35 sporocyst individual cells, focusing primarily on proliferating stem/germinal cells¹⁸ but most other cell types of this lifecycle stage remain uncharacterised. In the present study, we have followed an untargeted approach to characterise the individual transcriptomic signatures of more than 600 cells isolated from cultured mother sporocysts. Each of the tissue types were spatially validated by fluorescence in situ hybridisation (FISH). Furthermore, we have explored aspects of gene expression regulation, including the prediction of promoter motifs as tentative binding sites for transcription factors, in the stem/germinal and tegumental cell populations. This study contributes to the expansion of the currently scarce number single cell datasets for schistosomes⁵ and reveals key candidate genes involved in the intra-snail developmental phase of this neglected tropical disease pathogen.

Results

Six cell populations identified in the *Schistosoma mansoni* mother sporocyst

Freshly collected *S. mansoni* miracidia were transferred into sporocyst media to induce transformation into mother sporocysts. Within the first ~ 16 h most parasites have shed the cilia plates, and their tegument has been remodelled⁸; however, parasite *in-vitro* development is not synchronous. Therefore, we decided to culture the mother sporocysts for 5 days (named ‘D5 sporocysts’) to facilitate the complete transformation of > 95% of the parasites^{10,21} (Fig. 1A and Supplementary Fig. S1). The D5 sporocysts were collected and processed following the dissociation protocol previously used for schistosomula¹⁶, and live cells enriched and quantified using Fluorescence-activated Cell Sorting (FACS). The droplet-based 10X Genomics Chromium platform was used to generate transcriptome-sequencing data from a total of 601 cells after applying quality-control filters. With the assistance of a high-resolution nuclei quantification protocol, based on a machine-learning imaging platform²², we estimated that a D5 sporocyst contains an average of 169 nuclei (n = 5; range: 112–254) (Supplementary Fig. S2). Therefore, the number of quality-controlled cells theoretically represents > 3.5 × coverage of all cells in a single D5 mother sporocyst.

Based on top markers identified using Seurat, annotation from schistosomula single cell data and genes curated from the literature, we identified six discrete cell populations (Supplementary Table S3); Tegument-1 (138 cells), Tegument-2 (66 cells), Muscle (189 cells), Stem/germinal (119 cells), Parenchyma (66 cells) and Neuron (23 cells) (Fig. 1B). To further explore the biological processes in which the cells of each cluster are involved, we examined over-represented Gene Ontology terms using TopGO (Fig. 1C). Within each cluster, there were clear examples of single genes with annotated roles that are statistically enriched due to their rarity in the genome (Supplementary Tables S4, S5). Although many of these conformed to expectations—for instance, ‘neuropeptide signalling’ is enriched in the Neuron cluster due to the expression of the known marker gene *7B2*—we focussed on over-represented annotation supported by multiple genes. In the case of the Neuron cluster, this highlighted ‘neurotransmitter transport’, as well as less expected terms ‘nucleosome assembly’ and ‘proton transmembrane transport’. In the Muscle cluster, ‘cytoskeletal organisation’ was enriched, and previously characterised serotonin receptors in schistosomes²³ showed muscle-specific expression (Supplementary Fig. S3). The Parenchyma cluster showed enrichment for several aspects of metabolism including, ‘iron homeostasis’, ‘amino acid metabolism’ and ‘glycolysis’; whereas the Stem/germinal cell cluster, unsurprisingly, was highly enriched for genes involved in DNA replication, ribogenesis and translation.

In both Tegument clusters, ‘purine ribonucleotide/side salvage’ was highly enriched, as was ‘microtubule-based process’ due to the expression of 6–8 dyneins and the dynein-domain protein SmTAL2 (Fig. 1C and Supplementary Tables S4, S5). To further explore differences between the two Tegument subclusters, we first identified differentially expressed genes between Tegument 1 and 2 (Supplementary Tables S6, S7), and then performed a GO term enrichment analysis for each of the sub-clusters. ‘Proteolysis’, ‘metalloendopeptidase activity’, and ‘metal ion binding’ were GO terms supported by at least 2 genes found enriched in Tegument 1 (Supplementary Figure S4A and Table S8). On the other hand, Tegument 2 showed a significant enrichment in the GO terms ‘translation’, ‘structural constituent of ribosome’, ‘ribosome’ and ‘large ribosomal subunit’, i.e.,

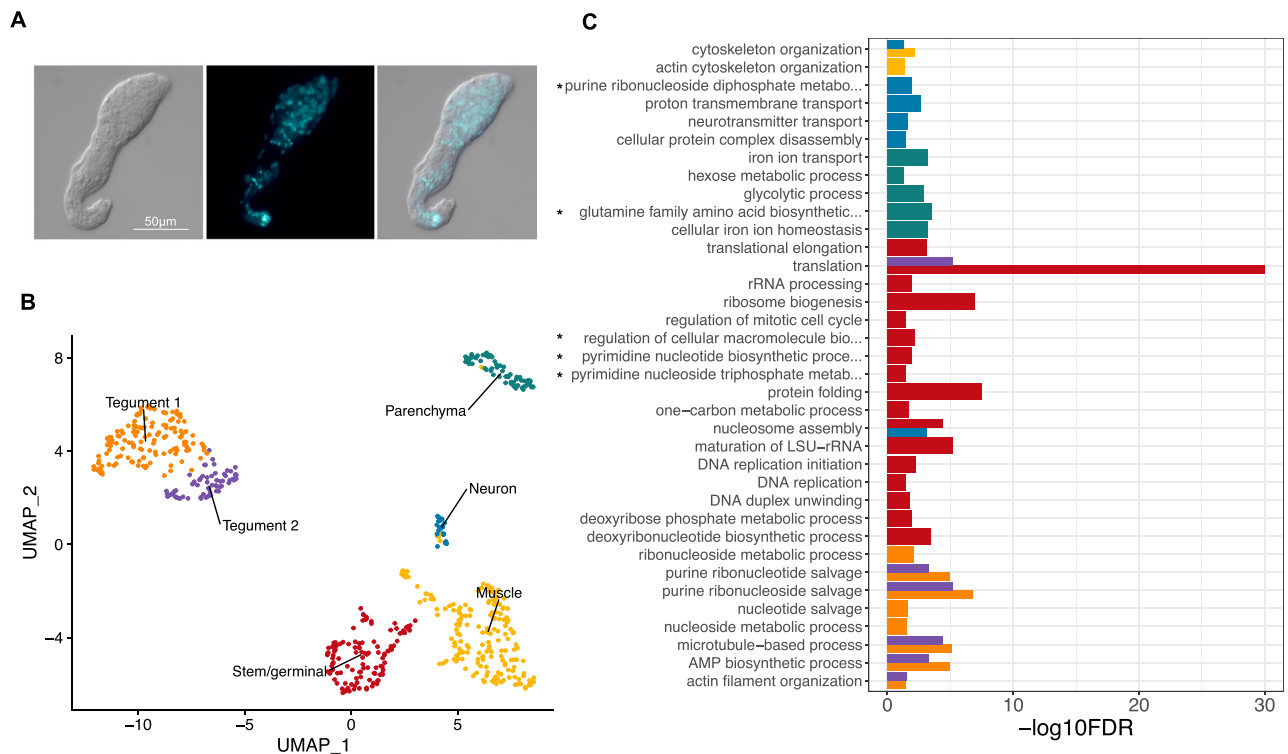


Figure 1. Six distinct cell populations identified in D5 mother sporocysts. **(A)** Representative picture of D5 sporocyst in DIC bright field (left), DAPI-fluorescent field (centre) and merged (right). Scale bar: 50 μm . **(B)** Uniform Manifold Approximation and Projection (UMAP) representation of 601 single cells from D5 sporocysts. The cell clusters are coloured and labelled as indicated. The list of the Seurat marker genes in all cell clusters is provided in Supplementary Table S3. **(C)** Gene ontology (GO) enrichment analysis for biological processes only (marker genes with minimum AUC=0.7, and GO terms supported by ≥ 2 genes), for top marker genes in each indicated colour-coded cell cluster (as shown in **B**). Only statistically significant GO enriched biological processes are depicted ($-\log_{10}(\text{FDR} < 0.05)$). FDR: False Discovery Rate. Full names of GO terms indicated with *: purine ribonucleoside diphosphate metabolic process (GO:0009179), glutamine family amino acid biosynthetic process (GO:0009084), regulation of cellular macromolecule biosynthetic process (GO:2000112), pyrimidine nucleotide biosynthetic process (GO:0006221), pyrimidine nucleoside triphosphate metabolic process (GO:0009147). Full data provided in Supplementary Tables S4 and S5, for non-specificity filtering analysis with AUC=0.7 and AUC=0.6, respectively.

biological processes and cellular components associated with protein synthesis. Remarkably, four out of the top 5 upregulated genes in Tegument 2 compared to Tegument 1 were ribosomal proteins (Supplementary Table S7); however, the proportion of cells in each of the subclusters expressing these genes was similar, e.g., the 40S ribosomal protein S24 is expressed in more than 90% of the cells in both Tegument subclusters (Supplementary Table S7). Most of the genes highly expressed in Tegument 2 were also expressed in Tegument 1, with few exceptions including an arrestin_C domain-containing protein (Smp_121950) expressed in ~ 54 and $\sim 74\%$ of the cells in Tegument 1 and 2, respectively (Supplementary Fig. S4B). On the contrary, several genes were enriched in Tegument 1, e.g. genes expressed in $> 90\%$ of Tegument 1 cells and $< 50\%$ of Tegument 2 cells, including a Plexin domain-containing protein (Smp_348500 -Supplementary Fig. S4B). Remarkably, 12 genes out of the top 15 upregulated genes in Tegument 1 are uncharacterised proteins (Supplementary Table S6).

The unexpected overrepresentation of biological processes associated with ‘nucleotide/side metabolism’ in the sporocyst tegument prompted us to compare among the most significantly represented GO terms and top marker genes in the tegument of other developmental stages for which single-cell transcriptomic data are available^{15–17}. The comparative analysis among the miracidium¹⁷, schistosomulum¹⁶, adult¹⁵ and sporocyst showed not only an expected GO term present across all stages (i.e., ‘microtubule-based process’), but also biological processes found only in the sporocyst tegument, including purine ‘ribonucleoside salvage’, ‘purine ribonucleotide salvage’, ‘nucleotide salvage’, biological processes linked to nucleic acid metabolism as indicated above (Supplementary Fig. S5 and Table S9). The comparison between top marker genes in the tegument cluster(s) across different developmental stages, revealed genes that may be specific to the sporocyst tegument with tentative functions associated with the absorption of molecules, e.g., Smp_329690 described as nose resistant to fluoxetine protein 6; Smp_169090, a major facilitator superfamily (MFS) domain-containing protein (Supplementary Fig. S6 and Table S10).

Following the comparative analysis across developmental stages performed for the tegument clusters and to further understand the similarities between life stages and the consistency of marker genes, we extended this analysis to other tissues. This revealed conserved and tentative developmental stage-specific GO terms and

top marker genes for stem cells (Supplementary Figs. S7, S8 and Tables S9, S11), parenchyma (Supplementary Figs. S9, S10 and Tables S9, S12), muscle cells (Supplementary Figs. S11, S12 and Tables S9, S13) and neurons (Supplementary Figs. S13, S14 and Tables S9, S14). It is important to highlight the limitations of this approach. The data obtained from each developmental stage were generated by different research teams, different versions of the genomes were employed for mapping, and different approaches were followed for the analyses (see Methods). Notwithstanding these caveats, several GO terms specific to the sporocyst tissues were identified, providing indications of tentative functional differences in this intra-molluscan stage compared to other developmental stages. For instance, the GO term ‘cellular protein complex disassembly’ is only found in neurons of sporocysts, but not in other developmental stages (Supplementary Table S9 and Fig. S13). Similarly, genes expressed in the sporocyst Parenchyma cluster were involved in carbohydrate, amino acid, lipid and iron metabolism, i.e., genes with tentative roles in the catabolism of nutrient molecules derived from the snail tissues. The enzymes UDP-glucose 4-epimerase (Smp_070780) and Ornithine aminotransferase (Smp_000660), the 14 kDa fatty acid-binding protein (Smp_095360), and Ferritin (Smp_063530, Smp_311630 and Smp_311640) were all within the top 30 marker gene list in the Parenchyma cluster of sporocysts, but not in the Parenchyma cluster(s) of the other developmental stages (Supplementary Figs. S9, S10 and Tables S9, S12).

Overall, these findings suggest that cluster-specific gene products display distinct functions in different molecular pathways, and that the identified cell populations may be involved in different biological processes in the mother sporocyst (Fig. 1C and Supplementary Tables S4 and S5). To spatially validate the predicted cell clusters, we defined highly specific cluster-defining marker genes (Fig. 2A and Supplementary Table S15), for which Fluorescence in situ Hybridization (FISH) probes were generated (Supplementary Table S2). We identified cells expressing the Muscle-specific marker myosin heavy chain, location of which correlated with actin filaments following the anterior–posterior axis of the sporocyst (Fig. 2B and Supplementary Fig. S15A, B). Whilst a Neuron-specific marker was expressed across a handful scattered cells in the mid region of the parasite (Fig. 2B and Supplementary Fig. S15C, D), cells expressing the Stem/germinal cluster-specific marker histone H2A were mainly located in clusters towards one pole (Fig. 2C). As stem cells are located in the posterior half of the miracidium larva¹⁷, we hypothesise this is the same in the mother sporocyst, so the Stem/germinal cell markers are thus likely to be highlighting the posterior end. In addition, a few individual Stem/germinal cells were located in the medial region towards the surface of the animal (Supplementary Fig. S15E, F). The two Tegument cell clusters that highly expressed the micro-exon gene 6 (*MEG-6*), were spatially validated by a strong FISH signal from cells lining the surface of the parasites (Fig. 2D, Supplementary Fig. S16 and Supplementary Video S1). The Parenchyma cluster cells were identified by the cluster specific expression of Smp_318890 (encoding a hypothetical protein). Even though this marker was expressed in <50% of the cells (Fig. 2A), it showed a high and specific expression in parenchyma cells. These cells seem to be distributed throughout the whole parasite body with a tendency towards the posterior pole (Fig. 2E, Supplementary Fig. S15G, H, and Supplementary Video S2). The parenchymal cells showed clear anterior–posterior cytoplasmic projections containing Smp_318890 transcripts (Fig. 2E, yellow arrowheads).

Stem cell heterogeneity revealed by self-assembling manifold algorithm

Using the self-assembling manifold (SAM) algorithm²⁴, the 119 stem/germinal cells were further analysed. Three discrete subclusters (clusters 0, 1 and 2) with distinct transcriptional profiles were identified (Figs. 3A, 3B and Supplementary Tables S16, S17). We used Scanpy²⁵ to rank genes that characterise each of the stem cell subclusters (Supplementary Table S18) and the five top-ranked genes for each cluster were used for visualisation. STRING²⁶ was used to predict protein–protein interactions among the top 50 ranked genes within each subcluster (Supplementary Table S19). The analysis of subcluster 0 showed a network with strong connectivity of inferred protein–protein interactions among the top marker genes (Figs. 3C). The top STRING terms for subcluster 0, across all categories, are related to ribosomes and translation (Supplementary Table S19). In contrast, weak connectivity was found between genes expressed in subclusters 1 and 2, with no network, far fewer STRING terms, and weaker statistical support (Supplementary Table S19).

Using TopGO we further explored the biological roles of cells within each stem/germinal subcluster by examining over-represented Gene Ontology terms. Related annotations from multiple genes within each cluster were evident; for example, and consistent with the STRING findings, subcluster 0 was highly enriched for biological processes ‘translation’ and ‘ribosome biogenesis’ (Figs. 3D and Supplementary Table S20). Statistically significant biological processes, supported by a minimum of two genes, were identified for the other two subclusters. In subcluster 1, ‘deoxyribonucleotide biosynthesis’ and ‘chromosome segregation’ suggest active DNA synthesis and cell division (Supplementary Table S20). In subcluster 2, ‘actin filament organisation’ and ‘microtubule-based process’, supported by three and six genes, respectively, as well as ‘purine ribonucleoside salvage’, ‘purine ribonucleotide salvage’ and ‘purine ribonucleoside monophosphate biosynthetic process’ indicate considerable similarity to the tegumental gene clusters described above (Supplementary Table S20).

Previously, three stem cell populations in mother sporocysts¹⁸ were identified as *kappa*, *delta* and *phi* based on the differential expression of seven marker genes (*klf*, *nanos-2*, *fgfrA*, *fgfrB*, *p53*, *zfp-1*, and *hes1*). Although these markers were expressed in our dataset, we were not able to use them to unambiguously assign the Stem/germinal cell subclusters identified by SAM to the *kappa*, *delta* and *phi* cell populations (Supplementary Figs. S17–S19).

Promoter motifs and transcription factor binding sites enriched in two major cell populations

To further investigate gene expression regulation in the Stem/germinal cell population, we searched for putative Transcription Factor Binding Sites (TFBSs). The Tegument-1 cell cluster was included as a control because it represents a somatic/differentiated cell population with a similar number of cells as the Stem/germinal cell population. Motif analysis was performed for marker genes specific to the Stem/germinal cluster (12 genes;

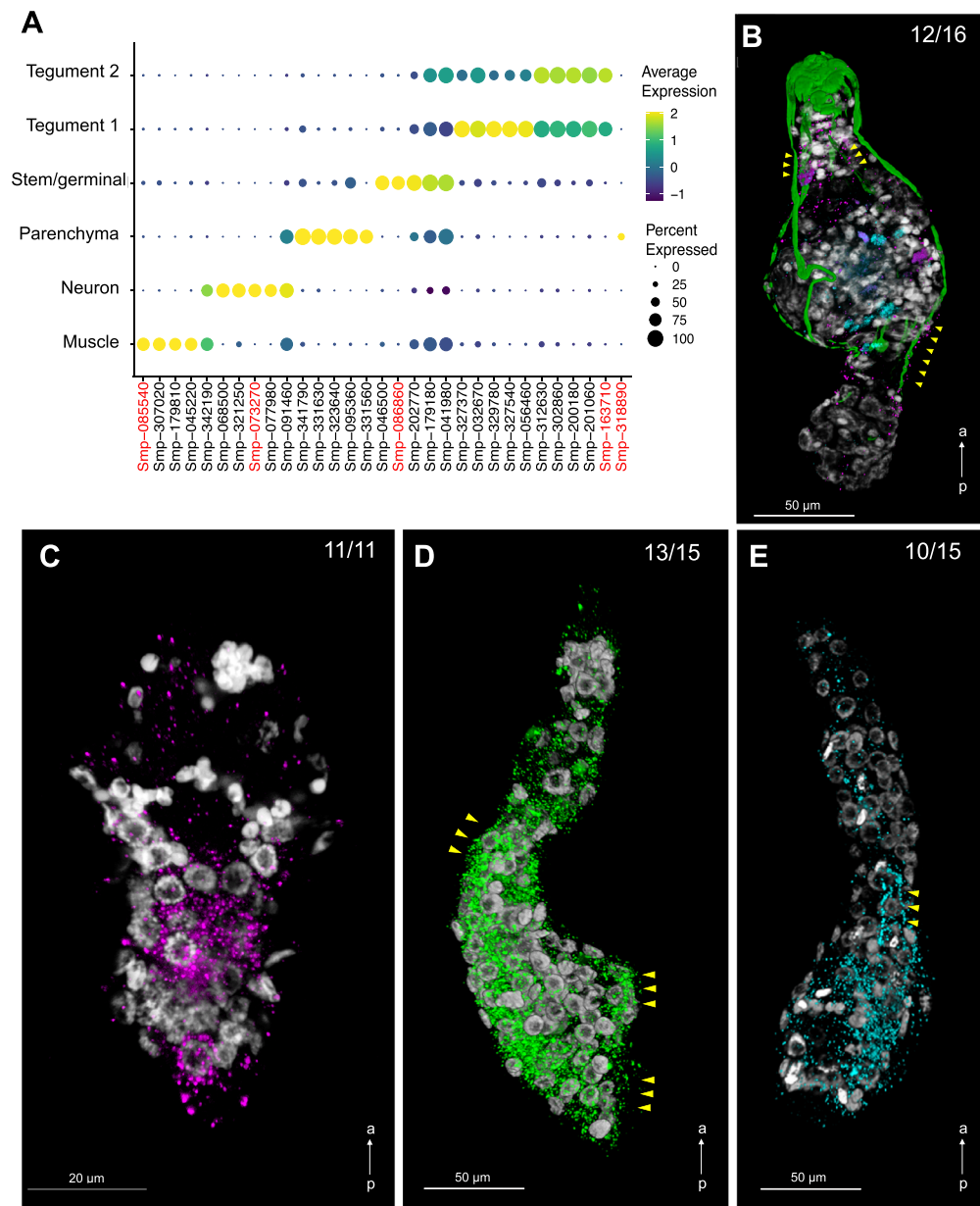


Figure 2. Cell clusters spatially validated by fluorescence in situ hybridization (FISH). **(A)** Dot plot showing the expression of the top 5 markers identified for each cell cluster. The average gene expression level for each marker is represented by a colour gradient from dark blue (low expression) to bright yellow (high expression). The circle sizes indicate the percentage of cells in each indicated cluster. The top marker genes for each cluster were defined as the highest AUC scores, which are calculated using FindAllMarkers (Seurat) using both presence and absence, and the level of expression. FISH probes for the following cluster-specific markers (highlighted in red) were used for spatial validation: Pan-tegument, *micro-exon gene 6* or *MEG-6* (Smp_163710); Muscle, *myosin heavy chain* (Smp_085540); Stem/Germinal, *histone H2A* (Smp_086860); Parenchyma, hypothetical protein (Smp_318890); and Neuron, *neuroendocrine protein 7b2* (Smp_073270). Full data for the indicated top markers are provided in Supplementary Table S15. **(B)** Double FISH with *myosin heavy chain* (Smp_085540; magenta) and *neuroendocrine protein 7b2* (Smp_073270; cyan) probes identified muscle and neuron cell clusters, respectively. Phalloidin-stained actin filaments are shown in green and DAPI staining in grey. Yellow arrowheads indicate co-localisation of *myosin heavy chain* and actin filaments. $n = 16$ parasites. **(C)** Localisation of the stem/germinal cells using FISH with *histone H2A* (Smp_086860; magenta). DAPI staining in grey. $n = 11$ parasites. **(D)** Localisation of tegumental cells using FISH with *MEG-6* (Smp_163710; green). DAPI staining in grey. Yellow arrowheads indicate cells lining the surface of the parasites. $n = 15$ parasites. **(E)** Localisation of parenchyma cells using FISH with *hypothetical protein* (Smp_318890; cyan). DAPI staining in grey. Yellow arrowheads point to cytoplasmic projections containing Smp_318890 transcripts. $n = 15$ parasites. Scale bars: 50 μ m in panels **(B)**, **(D)**, **(E)**, and 20 μ m in panel **(C)**. a \leftarrow p: anterior–posterior axis.

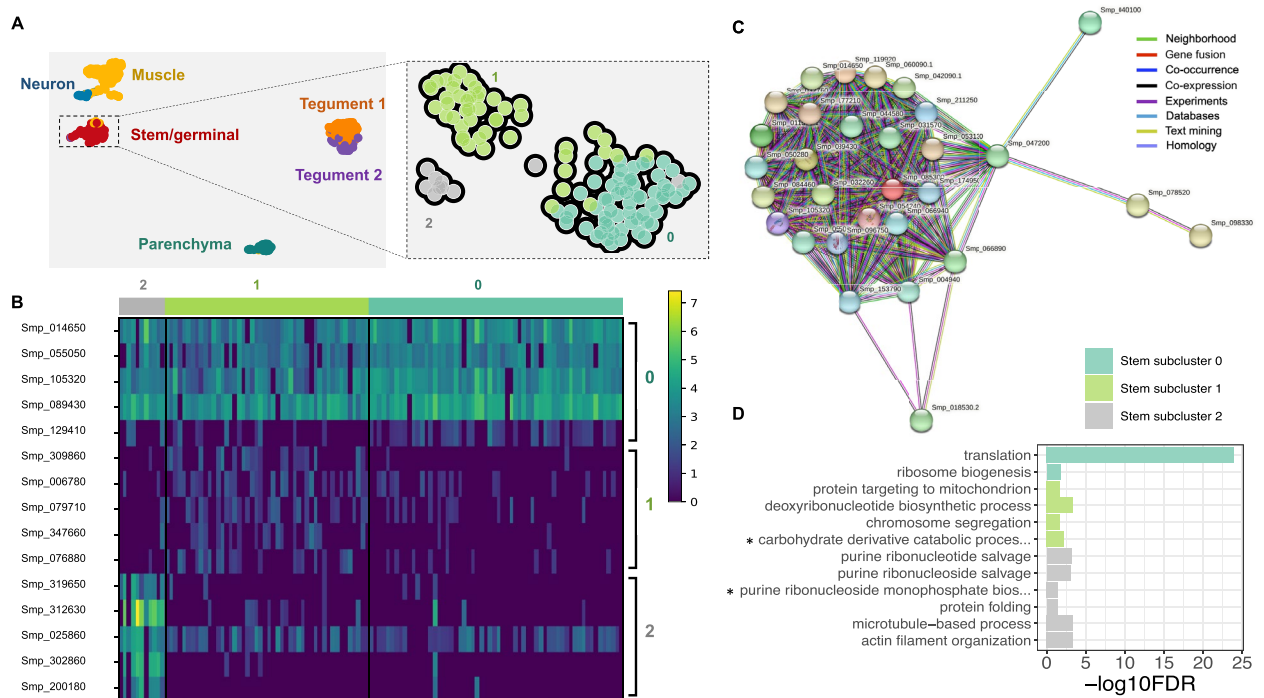


Figure 3. Stem/germinal cell sub-clusters. (A) Clustering of the sporocyst data using the self-assembling manifold (SAM) algorithm for all cells (left) and for the Stem/germinal cluster only (right). The SAM algorithm with Leiden clustering identified three stem/germinal sub-clusters (0, 1 and 2). The lists of SAM topology genes for all cell clusters or stem/germinal cell cluster only are provided in Supplementary Tables S16 and S17, respectively. (B) Heatmap of expression of the top 5 marker genes identified in each of the three stem/germinal sub-clusters. The average gene expression level for each marker is represented by a colour gradient from dark blue (low expression) to bright yellow (high expression). The full list of top marker genes identified by Scanpy in the three SAM stem cell sub-clusters is provided in Supplementary Table S18. (C) Interaction network analysis by STRINGdb for stem/germinal sub-cluster 0. The coloured nodes of the network represent proteins (genes ID for each protein are indicated, and all splice isoforms or post-translational modifications for each protein are collapsed). The coloured edges indicate sources of the interaction evidence as described. The full list of all enriched String terms for the top 50 markers of each stem/germinal sub-cluster is provided in Supplementary Table S19. (D) Stem/germinal subcluster gene ontology (GO) enrichment analysis in the category biological process, for top marker genes in each indicated colour-coded cell cluster. Only statistically significant GO enriched biological processes are depicted ($-\log_{10}(\text{FDR} < 0.05)$), and only terms supported by ≥ 2 genes are shown. FDR: False discovery rate. Full names of GO terms indicated with *: carbohydrate derivative catabolic process (GO:1901136), purine ribonucleoside monophosphate biosynthetic process (GO:0009168). Full data provided in Supplementary Table S20.

Fig. 4A) and the Tegment-1 cluster (49 genes; Supplementary Fig. S20A). We identified five enriched motifs in the promoter region (i.e., 1 kb upstream the Transcription Start Site—TSS) of Stem/germinal cell marker genes, and ten in those of tegumental cells (Fig. 4B, Supplementary Fig. S20B and Table S21). Interestingly, nearly all enriched motifs had ≥ 1 significant match(s) to known TFBSs in the model worm *C. elegans* based on JASPAR 2022 database, and no overlapping motifs were identified between the two analysed cell clusters (Supplementary Table S22). Within both the Stem/germinal (Fig. 4C) and Tegment-1 cell clusters (Supplementary Fig. S20C), most marker genes have binding sites for multiple transcription factors (TFs). Notably, from the 12 Stem/germinal cell marker genes, 11 contain the motif S-STREME-1, that is similar to the binding site for *ceh-22*, a homeobox gene and orthologue of the human *NKX2-2* gene (Fig. 4D). A further eight genes share the binding motif for *pha-4* (S-STREME-5), a forkhead/winged helix factor, and eight genes share a motif (S-STREME-4) similar to the TFBSs for both *ceh-22* and *vab-7* homeobox genes (Fig. 4C). Six genes share S-STREME-2, which is similar to the *unc-30* homeobox binding motif, and three genes share S-STREME-3, similar to the binding motif for the transcription factor *sma-4* (SMAD/NF-1 DNA-binding domain factors; Supplementary Table S22). Detailed information of motif locations within the promoter region of Stem/germinal and Tegment 1 marker genes is provided in Supplementary Table S23.

Using the above TFs from *C. elegans*, we identified putative *S. mansoni* orthologs. The putative *S. mansoni* TFs were examined by KEGG ortholog search, and the binding profiles were further confirmed for some candidates using the Jaspar Profile Inference tool²⁷. This screening produced a list of 12 candidate *S. mansoni* TF genes (Supplementary Table S24): one ortholog of *sma-4* (Smp_033950); two orthologs of *ceh-22* (Smp_027990 and Smp_186930), *pha-4* (FoxA) (Smp_331700 and Smp_158750) and *unc-30* (Smp_124010 and Smp_163140); and five orthologs of *vab-7* (Smp_147640, Smp_138140, Smp_347890, Smp_308310, Smp_134690). Two of the

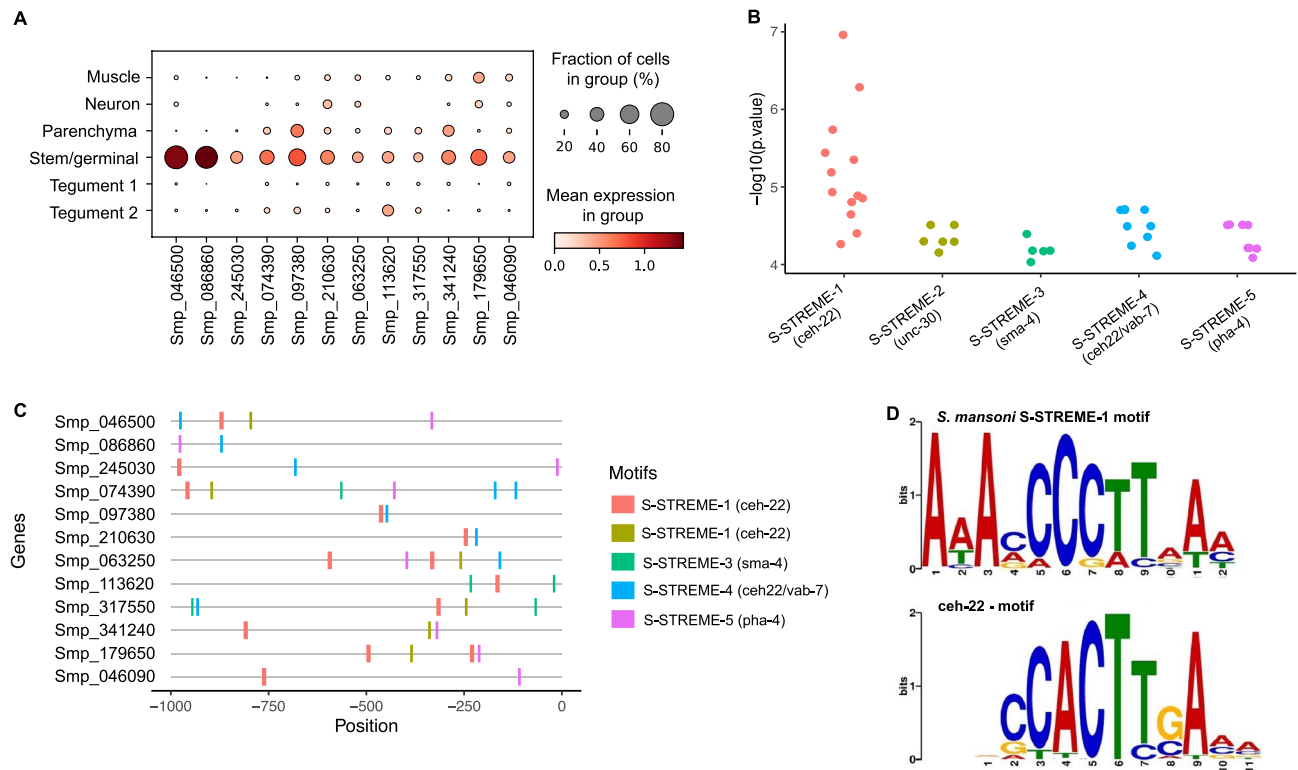
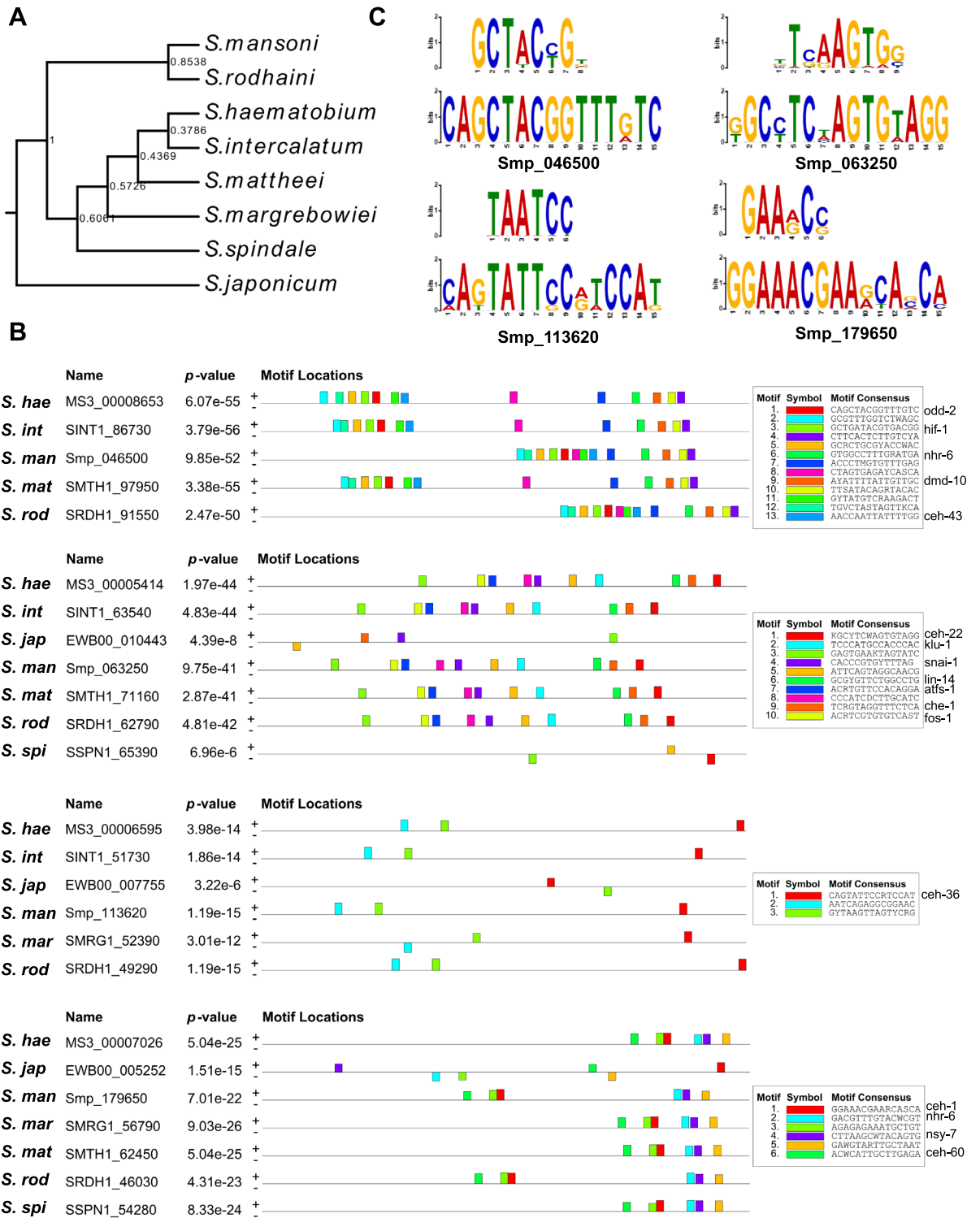


Figure 4. Promoter motif and transcription factor binding sites in stem/germinal cells. **(A)** Dot plot showing the expression level of the 12 Stem/germinal cell cluster-specific marker genes used for the analysis. Fraction of cells (%) and mean expression are indicated. The average gene expression level for each marker is represented by a colour gradient from white (low expression) to dark red (high expression). **(B)** Distribution of the $-\log_{10}(p \text{ values})$ for the top 5 ranked motifs identified in the 12 stem/germinal cell marker genes. The x-axis indicates motif names from XSTREME and significant match ($p < 0.05$) to known Transcription Factors Binding Sites (TFBSs) in the JASPAR 2022 nematode dataset (<https://jaspar.genereg.net/downloads/>). The y-axis represents log-transformed p values of each motif site shown in C. Full data provided in Supplementary Tables S21 and S22. **(C)** Predicted position distribution of the top 5 ranked motifs along the promoter region of the stem/germinal cell marker genes. The promoter region was taken as 1 kb upstream of the Transcription Start Site (TSS). Full data provided in Supplementary Table S23. **(D)** TF binding motifs found enriched in the promoter region of stem/germinal cell cluster marker genes. *Schistosoma mansoni* enriched motif named S-STREME-1 with the sequence: 1-AAAMCCCTTAAM (top) found in 11 of the 12 stem/germinal cell cluster marker genes with significant match to the binding site MA0264.1 (MA0264.1.ceh-22) for *C. elegans* ceh-22 (bottom) in the JASPAR database (<https://jaspar.genereg.net/>). The height of the letter in the motif scheme represents the frequency of the nucleotide observed in each indicated position.

12 candidates (Smp_186930 and Smp_033950) were not annotated as TFs in the KEGG search but indicated as orthologues to *C. elegans* counterparts. Based on a meta-analysis of gene expression for *S. mansoni* developmental stages²⁸, the Stem/germinal cell cluster specific markers (Fig. 4A) with promoter motifs and TFBSs identified above, tend to show a high expression in the ovary of mature females, miracidia and sporocysts (Supplementary Fig. S22A). Interestingly, the *sma-4* ortholog Smp_033950 (Mothers against decapentaplegic or MAD homolog) also showed a high expression across these developmental stages, particularly in sporocysts, relative to the other stages (Supplementary Figure S22A). At the single-cell level, this gene was expressed across all cell clusters, with the highest expression and percentage of cells in the Neuron cluster (Supplementary Figure S22B). Even though the TF Smp_033950 was expressed in $< 5\%$ of the stem/germinal cells (Supplementary Fig. S22B), it was the only one among the 12 putative TFs identified in this study (Supplementary Table S22) that showed a Stem/germinal cluster-specific expression and almost exclusively in subclusters 1 and 2 (Supplementary Fig. S22C).

To explore the evolutionary conservation of the predicted of TFBSs within the promoter region of Stem/germinal and Tegument cell cluster marker genes, we searched for enriched common regulatory binding sites across seven *Schistosoma* species in addition to *S. mansoni* (Fig. 5A and Supplementary Fig. S21A). Several conserved motifs and combinations of motifs were significantly enriched within 1 kb upstream of the TSS of *S. mansoni* marker gene orthologues across all the analysed *Schistosoma* species (Fig. 5B and C, Supplementary Fig. S21B and Supplementary Tables S25–S27). Moreover, the closer the species in the phylogenetic tree, the more conserved the sets of promoter motifs for each marker gene.

In addition to the predicted TFs and TFBSs expressed in Stem/germinal cluster and Tegument clusters, we investigated in our dataset the expression of well-characterised TFs in other developmental stages²⁹. The expression of the flatworm-specific zinc finger proteins genes *zfp-1* and *zfp-1-1*, involved in the tegument



◀ **Figure 5.** Promoter motif conservation of stem/germinal cell genes in the Schistosomatidae family. (A) Rooted species tree inferred with Orthofinder, showing the phylogenetic relationship between the indicated species. Branch support values are indicated. (B) Motifs found enriched in the promoter region of the orthologs in Schistosomatidae to the *S. mansoni* marker genes Smp_046500, Smp_063250, Smp_113620 and Smp_179650. The full list of analysed orthologous genes is provided in Supplementary Table S25. The colour-coded motifs detected for each group of orthologs are indicated along the 1 kb region upstream of the Transcription Start Site (TSS) for each gene. Significant matches with binding site found for *C. elegans* in the JASPAR database (<https://jaspar.genereg.net/>) were annotated. Full data provided in Supplementary Table S27. *S. hae*: *S. haematobium*, *S. int*: *S. intercalatum*, *S. man*: *S. mansoni*, *S. mat*: *S. mattheei*, *S. mar*: *S. margrebowiei*, *S. rod*: *S. rodhaini*, *S. spi*: *S. Spindale*, *S. jap*: *S. japonicum*. (C) Logo of most significant motifs identified for each group of promoter regions of orthologous genes and the corresponding binding site for *C. elegans* in the JASPAR database (<https://jaspar.genereg.net/>). Sequence CAGCTACGGTTTGTTC (bottom) found in orthologs of Smp_046500 showed a significant match to the binding site MA2148.1 (MA2148.1.odd-2) for *C. elegans* odd-2 (top); sequence KGCYTCWAGTGTAGG (bottom) found in orthologs of Smp_063250 showed a significant match to the binding site MA0264.2 (MA0264.2.ceh-22) for *C. elegans* ceh-22 (top); sequence CAGTATTCRTCCAT (bottom) found in orthologs of Smp_113620 showed a significant match to the binding site MA2159.1 (MA2159.1.ceh-36) for *C. elegans* ceh-36 (top); sequence GGAAACGAARCASCA (bottom) identified in orthologs of Smp_179650 showed a significant match to the binding site MA0260.1 (MA0260.1.che-1) for *C. elegans* che-1 (top). The height of the letter in the motif scheme represents the frequency of the nucleotide observed in each indicated position.

specification in adult worms, was detected in few sporocyst cells (<20) in the Stem/Germinal cluster (in the three subclusters—Supplementary Table S18) and in the Tegument (in particular in Tegument 2—Supplementary Table S3), respectively (Supplementary Figure S23).

Discussion

In this study, we employed single cell transcriptomics, followed by spatial validation, to unveil the cellular components of the mother sporocyst, the life cycle stage that marks the start of rapid asexual proliferation within the intermediate host. Due to the experimental challenges of studying parasites within snail tissues, we have used mother sporocysts cultured *in vitro* for 5 days, an approach that has been shown to mirror many aspects of *in-vivo* parasite development in snails^{30–33}. For instance, the first scRNA-seq study in schistosomes described the transcriptomic signatures of proliferating cells derived from mother sporocysts transformed *in vitro*¹⁸. The authors validated their findings in parasites developing *in vivo*, following parasite progeny through genesis of cercariae and intra-mammalian stages¹⁸. Molecular and cellular differences cannot be completely ruled out; however, *in-vivo* validation of observations across several stages of development¹⁸ provides strong evidence that the *in-vitro* miracidium–sporocyst transformation, and culture of early sporocyst stages, is a reliable model to study the parasite development.

Striking changes have been reported during the miracidium–sporocyst transition, including shedding of the ciliary plates, tegumental remodelling and an overall tissue reorganisation²¹ that transforms the free-living miracidium with ~365 cells¹⁷ into a simple ‘sac-like structure’ enriched with stem/germinal cells¹⁰. With the assistance of a novel machine learning approach²², we estimated an average of 169 nuclei in a single D5 mother sporocyst (ranging from 112 to 254 nuclei) and have revealed previously uncharacterised variability in cell number among the parasites at this developmental stage. The variability in cell number was expected, as asynchrony has been described during both *in-vitro* and *in-vivo* development of this and other developmental stages of schistosomes^{31,32,34}. The total number of sporocyst cells that passed the scRNA-seq QC-filtering steps was 601 (i.e., >3.5-fold the number of cells estimated to be present in a single D5 mother sporocyst). Similar to our previous findings in early schistosomula, using the same dissociation and cell staining protocol¹⁶, we were, however, not able to capture all cell types known to be present in this developmental stage, such as flame cells of the protonephridia system³⁵. Limitations have been previously reported for the FDA/PI staining of live/dead cells, and for the cell sorting, including unspecific FDA signals in the cell suspension³⁶, and fragile cells being destroyed or lost during the process³⁷, respectively. In addition, cell loss is unavoidable at each step of the protocol from parasite collection and dissociation into individual cells, to the 10X GEM Generation & Barcoding, cDNA purification, library preparation and sequencing. Optimised protocols to avoid dissociation-induced artefacts and to improve live/dead cell staining and cell sorting are being currently tested by us¹⁷ and others³⁸.

Notwithstanding the above limitations, we were able to capture from whole mother sporocysts the transcriptomic signatures of most of the cell types previously identified by electron microscopy in the parasite up to ten days post-infection⁶. We identified six discrete cell clusters that showed distinct transcriptomic signatures, with consistent enriched GO terms, and defined anatomic localisations in the whole animal, confirming previous electron microscopy findings^{6,8,10,11}. Interestingly, D5 sporocysts no longer show a well-defined brain structure, as described in the miracidium^{17,39}. The complex nervous system described for miracidium at the single cell level, comprising >45% of cells¹⁷, appears to have degenerated into only a handful of cells in the D5 sporocyst. Approximately 24 h after infection, mother sporocysts newly transformed *in vivo* already show neurons with evident degenerative changes¹⁰. The Muscle cluster marker *myosin heavy chain* was used to localise the myocytes that, as expected, colocalized with F-actin revealed by phalloidin stain. Muscle fibres also showed signs of degeneration, becoming thinner over time after the infection¹⁰. In spite of these degenerative changes in the neuromuscular system, sporocysts in culture are slightly motile and this activity increases when exposed to serotonin and serotonin receptor antagonists⁴⁰. We therefore analysed the expression of two previously characterised

serotonin receptors in intra-mammalian developmental stages²³, and both of them showed specific expression in the Muscle cluster. It has been reported that the infection of *Biomphalaria glabrata* snails with *S. mansoni* is associated with a reduction of serotonin and dopamine levels in the host brain⁴¹; however, the physiological role of these monoamine neurotransmitters on the sporocyst and its interaction with the host remains unknown.

The parenchyma cells identified in the D5 sporocyst showed a similar distribution to those in the miracidium¹⁷. Whilst these cells deliver energy, nutrients, and vitamins to the highly proliferating stem/germinal cells, they also start to degenerate and shrink to provide space for the early daughter sporocyst embryos^{10,21}. Moreover, particles of complex carbohydrates and lipid droplets have previously been described in the cytoplasm of parenchyma cells (historically described as ‘interstitial cells’)¹⁰. Indeed, it has been suggested that the parenchyma cells in the mother sporocyst store metabolic energy for use by the highly proliferative stem/germinal cells during the early development of daughter sporocyst embryos^{6,10}. This hypothesis is supported by the observation of parenchyma cells degenerating and dying while daughter sporocysts developed within the brood chamber in the mother sporocyst⁶. Consistent with this hypothesis, we identified key enzymes arguably involved in the catabolism of nutrient molecules derived from snail tissues (reviewed in⁴²), including UDP-glucose 4-epimerase and Ornithine aminotransferase, knock down of which by RNAi led to a significant reduction in cell proliferation in adult worms⁴³. Further experimental evidence is needed to test this hypothesis in intra-snail developmental stages.

The tegument of the miracidium is remodelled within the first ~ 5 h of infecting a snail, a process that involves shedding the ciliary plates and expanding the syncytial ridges between them to cover the parasite surface¹¹. Similar to the intra-mammalian developmental stages, the sporocyst tegument is a syncytium connected to cell bodies below the muscle layer, which has prompted comparisons of miracidium-to-sporocyst and cercaria-to-schistosomulum transformation mechanisms¹¹. Our FISH experiments confirmed that the tegument marker *MEG-6* is expressed on the surface of the parasite. The GO term analysis revealed unexpected evidence of high nucleotide/nucleotide metabolic activity in the tegument cells, including evidence of purine and pyrimidine salvage. These biological processes enriched in the sporocyst tegument may be specific to this developmental stage as none of these GO terms have been identified in the tegument of the miracidium¹⁷, schistosomulum¹⁶ or adult¹⁵. In fact, the GO term ‘microtubule-based process’ was the only biological process in the tegument shared among the four analysed developmental stages. However, and particularly due to the limitations of this comparative analysis elaborated in Methods, the presence of similar or related pathways cannot be ruled out in other developmental stages. For instance, ectonucleotidases that may be involved in purinergic catabolism have been described in the tegument of intra-mammalian stages⁴⁴. GO terms associated with microvilli (‘actin filament binding’ and ‘microtubule-based process’) were significantly enriched in the sporocyst tegument. By day 3 post infection, the sporocyst is completely covered with numerous microvilli that increase the exchange surface of the tegument revealing high absorptive activities during the rapid development of the parasite^{10,21}. Unlike the schistosome intra-mammalian developmental stages, the sporocysts lack a specialised digestive epithelium (i.e., gastrodermis). However, and similar to cestodes⁴⁵, our findings suggest that the absorption and metabolism of snail-derived nutrients may be accomplished by the sporocyst tegument. Consistent with this hypothesis is the presence in the tegument of the sporocyst, not in other stages, of the marker genes nose resistant to fluoxetine protein 6 (Smp_329690) and major facilitator superfamily (MFS) domain-containing protein (Smp_169090). The former is a *C. elegans* ortholog gene expressed in the intestine that binds and transports large hydrophobic molecules such as lipids and cholesterol⁴⁶. The latter, includes a MSF domain which is present in transmembrane transporters and binds to different substrates, including ions, carbohydrates, lipids, amino acids and peptides, nucleosides and other small molecules in both directions across the cell membrane⁴⁷. Further studies are needed to elucidate the biological meaning of these findings.

Schistosomiasis is a disease of stem cells, as these cells are the ultimate driver of life cycle propagation, development, and tissue homeostasis⁴⁸. The stem cell system has been studied by both ‘bulk’²⁰ and single cell-transcriptomic^{13,19} approaches in schistosomes and their free-living relatives, planaria⁴³. Three discrete stem cell populations, named *kappa*, *delta* and *phi* (based on their key marker genes) have previously been identified in the mother sporocyst^{4,18}. Furthermore, stem cell progenies were followed in vivo throughout the intra-snail and intra-mammalian developmental stages¹⁹. We used SAM, an algorithm that can find meaningful differences in systems with high intrinsic variability²⁴, to explore the presence of the three stem cell populations previously reported¹⁸. Even though three discrete Stem/germinal cell subclusters (named 0, 1 and 2) were identified, we were unable to unambiguously assign any of them to either *kappa*, *delta* or *phi* cell populations. Substantial technical differences between the two studies may explain these apparent discrepancies. In the present study, we used a dissociation protocol that enriched for all live cells and used 10X Genomics Chromium to capture single cells before sequencing. Wang et al.¹⁸ performed different dissociation—brief detergent treatment followed by trypsin digestion—then enriched EdU + proliferative stem cells by FACS and captured them individually on a microfluidic RNAseq chip before sequencing. Although the stem-cell-targeted study only analysed 35 individual cells, it enabled deeper sequencing per cell, hence higher resolution to accurately discriminate *kappa*, *delta* and *phi* cells¹⁸. Sub-cluster 0, identified in our dataset, showed a significant enrichment for GO terms associated with ribosome activity, mRNA processing and protein translation. This is consistent with the observation of germ cell ribosomes being replaced by polysomes one day after infection, and suggests that the cells are preparing for division¹⁰. Based on expression of the discriminatory marker *nanos* in subclusters 0 and 1, both clusters have some semblance of being *kappa* and *delta* cells, whereas subcluster 2 may correspond to the *phi* cell population because it lacked *nanos* expression and was enriched with genes commonly associated with the tegument such as dynein light chain and tetraspanin. Five days after infection, daughter sporocyst embryos begin to be surrounded by a primitive epithelium⁶. It is therefore tempting to speculate that the Stem/germinal subcluster 2 cells may be associated with somatic/temporary larva structures such as daughter sporocyst primitive epithelium, as has previously been described for *phi* cells^{4,18}. Further studies, including a temporal cell atlas of sporocysts⁵, and

in situ hybridisation and gene silencing of stem/germinal marker genes during the intra-snail development of the parasite¹⁸ will contribute to clarify this point.

Single cell transcriptomics have revealed tissue- and cell- specific regulatory networks, transcription factors (TFs) and transcription factor binding sites (TFBSs) in model organisms⁴⁹. Therefore, we decided to explore the enrichment of promoter motifs in marker genes specific for two key clusters: the Stem/germinal and Tegument-1 clusters. Furthermore, we extended the TFBSs prediction to seven additional *Schistosoma* species for which complete genomes and 5-UTR for ortholog marker genes are available^{50,51}. Notably, several motifs and combination of motifs, as well as distance between motifs were found significantly enriched and conserved within the marker gene promoter region across all the analysed species. The degree of conservation among the promoter motifs and sets of motifs correlated with the *Schistosoma* phylogeny. The evolutionary conservation of *cis*-regulatory sequences is striking throughout the evolution of mammalian genomes⁵², and a similar phenomenon may be true for parasitic flatworms. To the best of our knowledge, this has not yet been thoroughly investigated in parasitic flatworms warranting future studies.

The conservation of promoter motifs of specific marker genes across several species of *Schistosoma* strongly supports our TFBSs prediction. Most of the enriched promoter motifs predicted in *S. mansoni* Stem/germinal- and Tegument-1- specific marker genes showed similarity to known TFBSs in *C. elegans*, including sites for the homeobox genes *ceh-22*, *vab-7*, *unc-30*⁵³ and the forkhead/winged helix factor *pha-4*, ortholog of which (i.e. *S. mansoni foxA*) is expressed during the early development of the schistosomulum oesophageal gland⁵⁴, but not in the sporocyst. With the assistance of bulk RNAseq metadata across several developmental stages²⁸, we identified the schistosome orthologue of SMAD/NF-1 DNA-binding domain factor *sma-4* (Smp_033950, *Smad4* in WBPS18), highly expressed in ovary of mature females, miracidia and sporocyst and for which TFBSs were present in the promoter region of Stem/germinal cell cluster marker genes. The Smad proteins belong to a family of intracellular transducers in the Transforming Growth Factor β (TGF- β) pathway involved in *S. mansoni* development, host-parasite interaction and male–female pairing^{55,56}. *S. mansoni* Smad4 (SmSmad4) has previously been characterised⁵⁷, following the binding of TGF- β /Activin ligand to TGF- β receptor, SmSmad1 and SmSmad2 (receptor-regulated Smads or R-Smads) interact with SmSmad 4 in the cytoplasm, which in turn phosphorylate the Smad-complex that is translocated to the nucleus to control the expression of target genes⁵⁶. The relative expression of SmSmad1, 2 and 4 is stable across the intra-mammalian stages but, consistent with our own findings, SmSmad4 is overexpressed relative to the other two molecular partners in the egg and mother sporocyst⁵⁷. This would suggest a specific role during the development of intra-snail stages or interaction with the snail. The TGF- β pathway plays key roles in cell differentiation, proliferation and apoptosis, all critical processes for sporocyst development. Here, we have identified cell clusters where SmSmad 4 is expressed and its tentative promoter binding sites in target genes. However, future studies involving the analysis of the chromatin structure at the single cell level in tandem with scRNA-seq (e.g., single cell multimodal omics⁵⁸) will shine a light into the transcription factors associated with the TGF- β pathway.

In addition to the predicted TFBSs and TFs, we studied the expression of known TFs in other developmental stages. The flatworm-specific zinc finger proteins ZPF-1 (Smp_145470) and ZPF-1-1 (Smp_049580) have been described and well characterised by transcriptomic analysis, in situ hybridisation and RNAi experiments in *S. mansoni* adult worms²⁹. These TFs *zfp-1* and *zfp-1-1* are involved in the specification of tegumental progenitors, with *zfp-1* acting upstream to *zfp-1-1*. *zpf-1* is expressed in adult neoblast tegumental progenitors, whereas *zfp-1-1* is induced in cells that express tetraspanin 2 (*tsp-2*⁺ cells) and are committed towards tegument differentiation. These *tsp-2*/*zfp-1-1* positive cells extend cytoplasmic projections and eventually fuse with the tegumental syncytium²⁹. When analysing the expression of these TFs in our dataset; both *zfp-1* and *zfp-1-1* were expressed in few cells in the Stem/Germinal cluster (in the three subclusters) and only *zfp-1-1* in the Tegument, in particular in Tegument 2. As in the adult worms, in the mother sporocyst *zfp-1* was expressed in progenitor undifferentiated cells (Stem/germinal cell cluster), whereas *zfp-1-1* showed expression in the Tegument cluster, with few more cells in Tegument 2. This is consistent with the tegument specification described for adult worms suggesting that some pathways involved in cell commitment and differentiation may be conserved across developmental stages. Functional analysis to characterise developmental stage- and cell type- specific TFs are needed.

This study represents the first molecular characterisation of both somatic and germinal cells of the schistosome mother sporocyst. This developmental stage is essential for the parasite amplification that in turn facilitates a successful propagation of the life cycle and sustains the high prevalence of the infectious disease in endemic areas. A better understanding of mechanisms underlying these processes, including the discovery of drug targets involved in stem cell proliferation/ differentiation pathways conserved across the developmental stages may lead to novel control strategies for schistosomiasis and related trematodiasis.

Methods

Ethical statement

The complete life cycle of *Schistosoma mansoni* (NMRI strain) was maintained at the Wellcome Sanger Institute (WSI). All the animal regulated procedures were conducted under Home Office Project Licence No. P77E8A062 held by GR. All the protocols were presented and approved by the Animal Welfare and Ethical Review Body (AWERB) of the WSI. The AWERB is constituted as required by the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

Parasite material

On a monthly basis, TO outbred mice were infected as described⁵⁹. The animals were culled six weeks after infection and livers were removed and processed for egg isolation^{60,61}. Parasite material was prepared as previously described⁶¹. In brief, the livers were finely minced and digested overnight at 37 °C with gentle agitation in 1 × PBS

containing 0.5% of *Clostridium histolyticum* collagenase (Sigma), 200 U/ml penicillin, 200 µg/ml streptomycin and 500 ng/ml amphotericin B (ThermoFisher Scientific). The resulting mixture was filtered through sterile 250 µm and 150 µm sieves and subjected to a Percoll-sucrose gradient to purify the eggs, which were further washed three times in 1 × PBS and 200 U/ml penicillin, 200 µg/ml streptomycin and 500 ng/ml amphotericin B. The eggs were transferred to sterile water to induce hatching under light for at least 3 h, collecting miracidia and replacing the water every ~ 30 min. The miracidia were incubated for 30 min on ice, centrifuged at 800 g, 4 °C for 15 min, resuspended in sporocyst medium (MEMSE-J, 10% Fetal Bovine Serum, 10 mM Hepes, 100 U/ml penicillin, 100 µg/ml streptomycin—all reagents obtained from ThermoFisher Scientific) and cultured in a hypoxia chamber, with a gas mixture of 1% O₂, 3% CO₂ and 96% N₂, at 28 °C for 5 days^{60,61}.

Sporocyst cell number estimation by nuclear segmentation

In-vitro transformed mother sporocysts were cultured for 5 days (D5 sporocysts) and collected, washed three times by centrifugation, 400 g for 5 min, 1 × PBS and fixed overnight in 4% methanol-free paraformaldehyde (Pierce™) in 1 × PBS at 4 °C. The parasites were washed three times in 1 × PBS as above, resuspended in mounting media containing 4',6'-diamidino-2-phenylindole (DAPI) (Fluoromount-G™ Mounting Medium, with DAPI, Invitrogen), incubated overnight at 4 °C, and mounted on microscope slides. Z-stacks and maximum intensity projections were taken with a Leica SP8 confocal microscope. Nuclear segmentation was performed using the machine learning platform Ilastik (version 1.3.3)²². To enable the detection of nuclei with heterogeneous DAPI staining, pre-processing steps were first performed in ImageJ⁶², which involved applying image down-scaling, a Gaussian filter (2 pixel radius), and a median filter (2 pixel radius). A binary pixel classification was then trained in Ilastik²² using the auto-context pipeline, which includes two rounds of pixel classification. Training was performed on two small regions of interest from the dataset, representative of the total diversity of nuclear staining quality and intensity. Hysteresis thresholding was then used to segment the binary pixel classification map into distinct objects representing individual nuclei, with a core threshold of 0.93 and final threshold of 0.5. To further improve nuclear segmentation, thresholded images were imported to ImageJ using the Ilastik plugin, and a 3D distance-transform watershed was applied from the MorphoLibJ package⁶³.

Single-cell tissue dissociation and fluorescence-activated cell sorting (FACS)

D5 sporocysts were collected and processed for tissue dissociation as described¹⁶. Briefly, ~ 5000 D5 sporocysts were collected in 15-ml tubes and digested for 30 min in an Innova 4430 incubator with agitation at 300 rpm at 37 °C, using a digestion solution of 750 µg/ml Liberase DL (Roche) in 1 × PBS supplemented with 20% heat inactivated FBS. The resulting cell suspension was successively passed through 70 µm and 40 µm cells strainers, centrifuged at 300 g for 5 min and resuspended in cold 1 × PBS supplemented with 20% FBS. The cells were co-stained with 0.5 µg/ml of Fluorescein Diacetate (FDA; Sigma) and 1 µg/ml of Propidium Iodide (PI; Sigma) to label live and dead/dying cells, respectively. Live cells were enriched using BD Influx™ cell sorter (Becton Dickinson, NJ). It took 2–3 h from the enzymatic digestion to the generation of single-cell suspensions ready for loading on the 10X Genomics Chromium platform.

10X Genomics library preparation and sequencing

Suspensions of ~ 500 cells/µl live single-cells from mother sporocysts were loaded according to the standard protocol of the Chromium single-cell 3' kit to capture ~ 7000 cells per reaction (V2 chemistry). Thereafter, single-cell libraries were prepared and sequenced on an Illumina HiSeq4000 (custom read length: 26 bp read 1, 98 bp read 2, 8 bp index 1, 0 bp index 2 using the 75 bp PE kit), one sequencing lane per sample. Four biological replicates, i.e., 4 independent sporocyst batches generated from eggs collected in different occasions, were sequenced, and analysed. All raw sequence data is deposited in the ENA under the project accession ERP137194 and sample accession numbers: ERS11891013, ERS11891016, ERS11891015, ERS11891014 (Supplementary Table S1).

Mapping and single-cell RNA-seq quantification

The reference FASTA sequence for *S. mansoni* genome (version 9) was downloaded from WormBase Parasite release 17^{50,51} and gene annotations obtained from an intermediate version 9 stage before the WBPS release 17 (gtf accessible upon request from the authors). We followed previously described protocols with minor modifications^{16,17}. The scRNA-seq data were mapped to the *S. mansoni* reference genome using Cell Ranger (v 6.0.1), with default parameters. Approximately 62.2% of sequenced reads (average over four samples) mapped confidently to the transcriptome with a median UMI count per cell ranging from 973 to 1198 and a total of 857 cells detected (Supplementary Table S1).

Quality control, cell clustering and marker identification

The Seurat package (version 4.1.1)⁶⁴ (<https://satijalab.org/seurat/>) was used to analyse the filtered counts matrix produced by Cell Ranger. A maximum mitochondrial read threshold was set at 3% per cell. Analysis and quality control (QC) was performed according to the standard Seurat workflow using LogNorm (<https://satijalab.org/seurat/>)¹⁵. Following QC, 601 cells remained for downstream analyses. A combination of Seurat's JackStraw and Elbowplot were used to select the first 12 principal components for clustering cells. Clusters were generated using FindNeighbors and FindClusters, and clustree was used to choose the resolution for stable clustering. We manually segmented one cluster into two based on clear expression of neural markers, resulting in the identification of six cell clusters in total. Top markers for each cell cluster were identified using Seurat's FindAllMarkers function (test.use = "roc", only.pos = TRUE, return.thresh = 0), projecting cluster annotation from schistosomula scRNA-seq data¹⁶, and comparing to top markers curated from the literature^{4,18}. Seurat's

FindMarkers function (min.pct = 0.0, logfc.threshold = 0.0) was used to detect differentially expressed genes between the Tegument 1 and Tegument 2 clusters.

The data were further analysed using the self-assembling manifold (SAM) algorithm²⁴ (available at <https://github.com/atarashansky/self-assembling-manifold>) to assess cell cluster structure. The raw count data from the cells passing QC were exported from the Seurat object and imported to a SAM object, pre-processed and the SAM algorithm run using the default parameters (using Python version 3.8.5). To analyse the stem cells specifically, the above was repeated but only exporting cells from the annotated stem cell cluster (119 cells). After running the SAM algorithm to find topology, Leiden clustering was overlaid (param = 0.8). Scanpy²⁵ (version 1.8.2) was then used to extract summary QC information and the variable genes that characterised the three SAM clusters (rank_genes_groups). The cells were grouped by Leiden cluster to rank the genes that characterise each group (method = wilcoxon, corr.method = benjamini-hochberg).

Gene ontology (GO) and motif analysis

To create specific marker gene lists for Gene Ontology (GO) term analysis, the Seurat marker gene list was filtered for each cluster such that the remaining markers had an AUC score ≥ 0.7 (Supplementary Table S4). For the motif analyses (below), this gene list was further filtered for specificity, so each gene had to be detected in $\geq 70\%$ cells in the considered cluster, and $\leq 30\%$ of cells in the other clusters. GO analysis was also performed with a more permissive marker gene list filtering of AUC score ≥ 0.6 , and results are available in Supplementary Table S5.

GO enrichment analysis

GO term enrichment was performed using the weight01 method provided in topGO v2.46.0 (available at <http://bioconductor.org/packages/release/bioc/html/topGO.html>) for the Biological Process (BP) and Molecular Function (MF) aspects of the ontology. Analysis was restricted to terms with a node size of ≥ 5 . Fisher's exact test was applied to assess the significance of overrepresented terms compared with all expressed genes. The threshold was set as FDR < 0.05 . The top 50 ranked genes for each of the three Stem/germinal subclusters (produced using Scanpy) were used as the basis to calculate enriched GO terms in the subclusters, using this same method and parameters.

To characterise the differences between Tegument 1 and Tegument 2, the list of differentially expressed genes between these clusters (described above) was split into those with higher expression in Tegument 1, and Tegument 2, and were filtered to retain only those with an adjusted p value lower than 0.001 (Supplementary Tables S6, S7). The GO enrichment analysis was then performed using the same method as described here for all the clusters, but also including the Cellular Component (CC) aspect of the ontology (Supplementary Table S8).

Analysis of GO term enrichment for BP was also performed for comparison with other *S. mansoni* developmental stages for which single-cell transcriptomic data are available. GO term analysis tables from previously published developmental stages were extracted from supplementary tables^{15,17} or manually¹⁶. RVen (v1.1.0) was used to find and visualise overlapping GO terms (by GO ID) in sporocyst tegument clusters with tegument GO terms from miracidia¹⁷, schistosomula¹⁶, and adult worms¹⁵ and then expanded by completing this same analysis for all tissue types detected in the sporocyst (Supplementary Table S9).

Shared gene analysis across developmental stages

Marker genes, extracted from single cell transcriptomic published data for miracidia¹⁷, schistosomula¹⁶, and adult¹⁵ stages, were combined with the marker genes identified here. For each Stem/germinal, Muscle, Neuron, Tegument, and Parenchyma cluster(s) we selected the top marker genes from each cluster of that tissue type in each stage. Critically, this comparison is limited by three particular constraints; (1) the developmental stages were sequenced at four separate times in different laboratories, and data mapped to three different versions of the genome (schistosomula and adults to v7, sporocysts to v9, and miracidia to v10), hence, inconsistencies in gene annotations may be expected and IDs for some genes may be missed in some of the analysed stages; (2) the data were analysed following different methods, therefore approaches were varied and the same metrics were not available for all marker lists; (3) the marker gene lists are generally targeted to finding genes specific to clusters, rather than across tissues, thus, some pan-tissue markers may not be captured in the top 30 markers, especially for tissues with many subclusters (e.g. neurons). Therefore, shared markers are likely to be accurate, but where markers are not shared between stages it may be due to one or more of the above confounders. For miracidia, sporocysts, and schistosomula the markers were filtered to have a minimum AUC of 0.6, and for the adults the top 200 genes for each cluster by adjusted p value. We then selected the top 25 genes from each cluster, and from this list selected the top 30 genes across all clusters for the given tissue type in each stage. If there was only one cluster for a given tissue, then the top 30 genes were selected for that cluster. 'Top' was defined by log-fold change in miracidia, sporocysts, and adults, and average.diff in schistosomula. RVen (v1.1.0) was used to find and visualise common genes across the life stages for each of the above tissues (Supplementary Tables S10–S14).

Motif analysis in the promoter region of marker genes

The sequences of promoter regions (defined as 1 kb upstream of annotated transcriptional start sites or the intervening distance to the upstream gene if < 1 kb) of marker genes were extracted based on the *S. mansoni* v9 annotation and fed to the MEME Suite tool XSTREME (v5.4.1; <https://meme-suite.org/meme/tools/xstreme>)⁶⁵. We used shuffled input sequences with default Markov order as the control. In addition, the non-redundant transcription factor (TF) binding profiles (Position Frequency Matrix) for nematodes were obtained from the JASPAR 2022 database²⁷ and supplied to XSTREME. A default threshold E -value < 0.05 was applied to select enriched motifs in the promoter regions and E -value < 1 for matching the motif to JASPAR TF binding sites using the Tomtom tool. Based on the discovered transcription factor binding sites (TFBS), peptide sequences

of corresponding transcription factors in *C. elegans* were retrieved from WormBase⁶⁶ and aligned to *S. mansoni* proteins using BlastP with *E*-value < 1e-5; *C. elegans* was chosen as having a particularly well-curated set of TFBS and TFs. The top 10 *S. mansoni* protein hits for each *C. elegans* TF (including isoforms) were checked for their TF identity based on KEGG⁶⁷ ortholog search using the GHOSTX program and BBH method, and their binding profiles were validated again using the Jaspar Profile Inference tool (<https://jaspar.genereg.net/inference>).

To study the motif conservation of regulatory binding sites in the Schistosomidae family, the promoter region, i.e., 1 kb upstream the Transcription Starting Site (TSS) of orthologous genes to the ones predicted to be under regulation in *Schistosoma mansoni* were retrieved from seven *Schistosoma* species; *S. japonicum* (PRJNA520774), *S. haematobium* (PRJNA78265), *S. intercalatum* (TD1 PRJEB44434), *S. margrebowiei* (PRJEB44434), *S. mattheei* (PRJEB44434), *S. rodhaini* (TD1 PRJEB44434), and *S. spindale* (PRJEB44434) with BioMart (<https://parasite.wormbase.org/info/Tools/biomart.html>). Genes with no predicted 5'UTR or that showed paralogous expansions were excluded from the analysis.

Blastn with parameters *E*-value < 0.1 and word_size 5 was used to detect homologous motifs in the promoter regions of other *Schistosoma* species using the transcription factor binding sites motifs detected in the promoters of *S. mansoni* genes as query. Meme Suite tool XSTREME64 with the same parameters as described above, was used to predict conserved motifs in the promoter regions of orthologous genes in the Schistosomatidae family.

Orthofinder⁶⁸ with default parameters was used to infer a rooted phylogenetic tree based on the proteome of the *Schistosoma* species indicated above.

Protein–protein interaction network analysis

To predict protein–protein interactions within the stem/germinal subcluster top markers, we used the online tool STRING²⁶ (<https://string-db.org>). In STRINGdb, interaction networks were constructed using evidence sources as edges and a minimum interaction score of 0.9 for the top 50 marker genes from the SAM analysis. Data available in Supplementary Table S19.

Fluorescence in situ hybridisation

To spatially validate the main cell populations, we used fluorescence in situ hybridisation (FISH) following the third generation in situ hybridisation chain reaction (HCR) approach⁶⁹. Buffers, hairpins and probes against top marker gene transcripts for tegument (*micro-exon gene 6* or *MEG-6*; Smp_163710), muscle (*myosin heavy chain*; Smp_085540), stem/germinal cell (*histone H2A*; Smp_086860), parenchyma (*hypothetical protein*; Smp_318890) and neuronal (*neuroendocrine protein 7b2*; Smp_073270) cell clusters were purchased from Molecular Instruments (Los Angeles, California, USA) (Supplementary Table S2). D5 sporocysts were processed for HCR as described^{16,17,70}, with minor modifications. In brief, the D5 sporocysts were collected, washed once in sporocyst media (above) by centrifugation at 500 g, 5 min at room temperature (RT), and fixed in 4% PFA in PBSTw (1 × PBS + 0.1% Tween) at RT for 30 min with gentle rocking. The fixed sporocysts were transferred into an incubation basket (Intavis, 35 µm mesh) in a 24-well plate, rinsed 5 times for 5 min in ~750 µl of PBSTw, and incubated in 750 µl of 1 × PBSTw containing 2 µl of Proteinase K (20 mg/ml) (ThermoFisher Scientific) at RT with no agitation. Thereafter, the sporocysts were rinsed twice in ~750 µl of PBSTw for 5 min at RT, incubated in 500 µl of PBSTw with 0.2 mg/ml glycine on ice for 15 min and rinsed twice as above. Thereafter, the parasites were distributed into different baskets depending on the number of experimental groups to test (up to two probes were multiplexed per experimental group), and incubated in 750 µl of 1:1 PBSTw: hybridisation buffer for 5 min at RT. The 1:1 PBSTw: hybridisation buffer solution was replaced with 750 µl of hybridisation buffer, and parasites placed in a humidified chamber and incubated at 37 °C for 1 h (i.e., pre-hybridisation step). During the incubation, 1 to 2 µl of indicated HCR probe was added to 650 µl of hybridisation buffer and equilibrated to 37 °C. After the pre-hybridisation step, the hybridisation buffer was replaced with the hybridisation buffer containing the HCR probe(s), and parasites incubated at 37 °C in the humidified chamber overnight. The parasites were rinsed four times for 15 min in probe wash buffer, previously equilibrated at 37 °C, followed by four washes in filtered 5 × SSCTw (5 × SSC buffer + 0.1% Tween) at RT on gentle rocker and incubated in 600 µl of amplification buffer, previously equilibrated at RT, for 30 min at RT on gentle rocker. During this pre-amplification step, 10 µl of corresponding hairpin for each probe were incubated at 95 °C for 90 s, cooled to RT, in the dark for 30 min and added to 600 µl of amplification buffer. Thereafter, the amplification buffer was replaced with amplification buffer containing the hairpin mix, and the plate was incubated in the dark, at RT overnight. The larvae were rinsed in 5 × SSCTw, three times for 5 min followed by a long wash of 30–60 min on gentle rocker in the dark. In some experiments, the sporocysts were incubated in phalloidin (1 µl of phalloidin diluted in 2 ml of 5 × SSCTw) and incubated in the dark for ~60 min followed by three washes of 5 min each and four washes of 30 min each in 5 × SSCTw. After the washes, the parasites were incubated in mounting media containing DAPI (Fluoromount-G™ Mounting Medium, with DAPI, Invitrogen), incubated overnight at 4 °C, and mounted on microscope slides. For each HCR experiment, controls of hairpin with no probe (i.e., control for signal specificity), and no probe no hairpin (i.e., control for autofluorescence) were included. Z-stacks and maximum intensity projections were taken with a Leica SP8 confocal microscope. FISH experiments were performed at least twice (≥ 2 different batches of sporocysts collected from different groups of infected mice) for each HCR probe. The number of screened parasites (*n*) are indicated in the Figure legends.

Data availability

All raw sequence data is deposited in the ENA under the project accession ERP137194 and sample accession numbers: ERS11891013, ERS11891016, ERS11891015, ERS11891014 (Supplementary Table S1). The scripts used to perform the analyses presented here are available at <https://github.com/tessatten/singlecell-sporocysts>. 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. Supplementary Information 27.

Received: 5 July 2023; Accepted: 27 February 2024

Published online: 12 March 2024

References

1. Ayabina, D. V. *et al.* Gender-related differences in prevalence, intensity and associated risk factors of *Schistosoma* infections in Africa: A systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* **15**, e0009083. <https://doi.org/10.1371/journal.pntd.0009083> (2021).
2. Lo, N. C. *et al.* Review of 2022 WHO guidelines on the control and elimination of schistosomiasis. *Lancet Infect. Dis.* **22**, e327–e335. [https://doi.org/10.1016/S1473-3099\(22\)00221-3](https://doi.org/10.1016/S1473-3099(22)00221-3) (2022).
3. Vale, N. *et al.* Praziquantel for schistosomiasis: Single-drug metabolism revisited, mode of action, and resistance. *Antimicrob. Agents Chemother.* <https://doi.org/10.1128/AAC.02582-16> (2017).
4. Nanes Sarfati, D., Li, P., Tarashansky, A. J. & Wang, B. Single-cell deconstruction of stem-cell-driven schistosome development. *Trends Parasitol.* **37**, 790–802. <https://doi.org/10.1016/j.pt.2021.03.005> (2021).
5. Rinaldi, G., Attenborough, T., Rawlinson, K. A. & Berriman, M. Schistosomes go single cell. *Nat. Rev. Microbiol.* **20**, 705. <https://doi.org/10.1038/s41579-022-00810-0> (2022).
6. Meuleman, E. A., Holzmann, P. J. & Peet, R. C. The development of daughter sporocysts inside the mother sporocyst of *Schistosoma mansoni* with special reference to the ultrastructure of the body wall. *Z. Parasitenkd.* **61**, 201–212. <https://doi.org/10.1007/BF00925512> (1980).
7. Cheng, T. C. & Bier, J. W. Studies on molluscan schistosomiasis: An analysis of the development of the cercaria of *Schistosoma mansoni*. *Parasitology* **64**, 129–141. <https://doi.org/10.1017/s003118200004470x> (1972).
8. Basch, P. F. & DiConza, J. J. The miracidium-sporocyst transition in *Schistosoma mansoni*: Surface changes in vitro with ultrastructural correlation. *J. Parasitol.* **60**, 935–941 (1974).
9. Koie, M. & Frandsen, F. Stereoscan observations of the miracidium and early sporocyst of *Schistosoma mansoni*. *Z. Parasitenkd.* **50**, 335–344. <https://doi.org/10.1007/BF02462978> (1976).
10. Pan, S. C. *Schistosoma mansoni*: the ultrastructure of larval morphogenesis in *Biomphalaria glabrata* and of associated host-parasite interactions. *Jpn. J. Med. Sci. Biol.* **49**, 129–149. <https://doi.org/10.7883/yoken1952.49.129> (1996).
11. Samuelson, J. C. & Caulfield, J. P. Role of pleated septate junctions in the epithelium of miracidia of *Schistosoma mansoni* during transformation to sporocysts in vitro. *Tissue Cell* **17**, 667–682. [https://doi.org/10.1016/0040-8166\(85\)90003-5](https://doi.org/10.1016/0040-8166(85)90003-5) (1985).
12. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-023-00580-2> (2023).
13. Kulkarni, A., Anderson, A. G., Merullo, D. P. & Konopka, G. Beyond bulk: A review of single cell transcriptomics methodologies and applications. *Curr. Opin. Biotechnol.* **58**, 129–136. <https://doi.org/10.1016/j.copbio.2019.03.001> (2019).
14. Aldridge, S. & Teichmann, S. A. Single cell transcriptomics comes of age. *Nat. Commun.* **11**, 4307. <https://doi.org/10.1038/s41467-020-18158-5> (2020).
15. Wendt, G. *et al.* A single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. *Science* **369**, 1644–1649. <https://doi.org/10.1126/science.abb7709> (2020).
16. Diaz Soria, C. L. *et al.* Single-cell atlas of the first intra-mammalian developmental stage of the human parasite *Schistosoma mansoni*. *Nat. Commun.* **11**, 6411. <https://doi.org/10.1038/s41467-020-20092-5> (2020).
17. Attenborough, T. *et al.* A single-cell atlas of the free-living miracidium larva of *Schistosoma mansoni*. *bioRxiv*. Doi: <https://doi.org/10.1101/2023.03.27.533868> (2023).
18. Wang, B. *et al.* Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *Elife* <https://doi.org/10.7554/eLife.35449> (2018).
19. Li, P. *et al.* Single-cell analysis of *Schistosoma mansoni* identifies a conserved genetic program controlling germline stem cell fate. *Nat. Commun.* **12**, 485. <https://doi.org/10.1038/s41467-020-20794-w> (2021).
20. Wang, B., Collins, J. J. 3rd. & Newmark, P. A. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *Elife* **2**, e00768. <https://doi.org/10.7554/eLife.00768> (2013).
21. Poteaux, P., Gourbal, B. & Duval, D. Time series analysis of tegument ultrastructure of in vitro transformed miracidium to mother sporocyst of the human parasite *Schistosoma mansoni*. *Acta Trop.* **240**, 106840. <https://doi.org/10.1016/j.actatropica.2023.106840> (2023).
22. Berg, S. *et al.* ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232. <https://doi.org/10.1038/s41592-019-0582-9> (2019).
23. Patocka, N., Sharma, N., Rashid, M. & Ribeiro, P. Serotonin signaling in *Schistosoma mansoni*: A serotonin-activated G protein-coupled receptor controls parasite movement. *PLoS Pathog.* **10**, e1003878. <https://doi.org/10.1371/journal.ppat.1003878> (2014).
24. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *Elife* <https://doi.org/10.7554/eLife.48994> (2019).
25. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15. <https://doi.org/10.1186/s13059-017-1382-0> (2018).
26. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612. <https://doi.org/10.1093/nar/gkaa1074> (2021).
27. Castro-Mondragon, J. A. *et al.* JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173. <https://doi.org/10.1093/nar/gkab1113> (2022).
28. Lu, Z., Zhang, Y. & Berriman, M. A web portal for gene expression across all life stages of *Schistosoma mansoni*. *bioRxiv*. <https://doi.org/10.1101/308213> (2018).
29. Wendt, G. R. *et al.* Flatworm-specific transcriptional regulators promote the specification of tegumental progenitors in *Schistosoma mansoni*. *Elife* <https://doi.org/10.7554/eLife.33221> (2018).
30. DiConza, J. J. & Basch, P. F. Axenic cultivation of *Schistosoma mansoni* daughter sporocysts. *J. Parasitol.* **60**, 757–763 (1974).
31. Ivanchenko, M. G. *et al.* Continuous in vitro propagation and differentiation of cultures of the intramolluscan stages of the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4965–4970. <https://doi.org/10.1073/pnas.96.9.4965> (1999).
32. Ouyi, Y. *et al.* *Schistosoma mansoni* larvae in vitro cultures using *Biomphalaria glabrata* extracts. *Acta Trop.* **235**, 106636. <https://doi.org/10.1016/j.actatropica.2022.106636> (2022).
33. Yoshino, T. P. & Laursen, J. R. Production of *Schistosoma mansoni* daughter sporocysts from mother sporocysts maintained in synxenic culture with *Biomphalaria glabrata* embryonic (Bge) cells. *J. Parasitol.* **81**, 714–722 (1995).
34. Wangwivatsin, A. *et al.* Transcriptome of the parasitic flatworm *Schistosoma mansoni* during intra-mammalian development. *PLoS Negl. Trop. Dis.* **14**, e0007743. <https://doi.org/10.1371/journal.pntd.0007743> (2020).

35. Collins, J. J. III., King, R. S., Cogswell, A., Williams, D. L. & Newmark, P. A. An atlas for *Schistosoma mansoni* organs and life-cycle stages using cell type-specific markers and confocal microscopy. *PLoS Negl. Trop. Dis.* **5**, e1009. <https://doi.org/10.1371/journal.pntd.0001009> (2011).
36. Boyd, V., Cholewa, O. M. & Papas, K. K. Limitations in the use of fluorescein diacetate/propidium iodide (FDA/PI) and cell permeable nucleic acid stains for viability measurements of isolated islets of langerhans. *Curr. Trends Biotechnol. Pharm.* **2**, 66–84 (2008).
37. Hu, P., Zhang, W., Xin, H. & Deng, G. Single cell isolation and analysis. *Front. Cell Dev. Biol.* **4**, 116. <https://doi.org/10.3389/fcell.2016.00116> (2016).
38. Machado, L., Relaix, F. & Mourikis, P. Stress relief: Emerging methods to mitigate dissociation-induced artefacts. *Trends Cell Biol.* **31**, 888–897. <https://doi.org/10.1016/j.tcb.2021.05.004> (2021).
39. Jurberg, A. D. *et al.* The embryonic development of *Schistosoma mansoni* eggs: Proposal for a new staging system. *Dev. Genes Evol.* **219**, 219–234. <https://doi.org/10.1007/s00427-009-0285-9> (2009).
40. Boyle, J. P., Zaide, J. V. & Yoshino, T. P. *Schistosoma mansoni*: Effects of serotonin and serotonin receptor antagonists on motility and length of primary sporocysts in vitro. *Exp. Parasitol.* **94**, 217–226. <https://doi.org/10.1006/expr.2000.4500> (2000).
41. Manger, P., Li, J., Christensen, B. M. & Yoshino, T. P. Biogenic monoamines in the freshwater snail, *Biomphalaria glabrata*: Influence of infection by the human blood fluke, *Schistosoma mansoni*. *Comp. Biochem. Physiol. A Physiol.* **114**, 227–234. [https://doi.org/10.1016/0300-9629\(95\)02131-0](https://doi.org/10.1016/0300-9629(95)02131-0) (1996).
42. El-Ansary, A. Biochemical and immunological adaptation in schistosome parasitism. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **136**, 227–243. [https://doi.org/10.1016/s1096-4959\(03\)00124-6](https://doi.org/10.1016/s1096-4959(03)00124-6) (2003).
43. Wang, J. *et al.* Large-scale RNAi screening uncovers therapeutic targets in the parasite *Schistosoma mansoni*. *Science* **369**, 1649–1653. <https://doi.org/10.1126/science.abb7699> (2020).
44. Skelly, P. J., Nation, C. S. & Da’Dara, A. A. *Schistosoma mansoni* and the purinergic halo. *Trends Parasitol.* **38**, 1080–1088. <https://doi.org/10.1016/j.pt.2022.09.001> (2022).
45. Tsai, I. J. *et al.* The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496**, 57–63. <https://doi.org/10.1038/nature12031> (2013).
46. Choy, R. K., Kemner, J. M. & Thomas, J. H. Fluoxetine-resistance genes in *Caenorhabditis elegans* function in the intestine and may act in drug transport. *Genetics* **172**, 885–892. <https://doi.org/10.1534/genetics.103.024869> (2006).
47. Madej, M. G., Dang, S., Yan, N. & Kaback, H. R. Evolutionary mix-and-match with MFS transporters. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5870–5874. <https://doi.org/10.1073/pnas.1303538110> (2013).
48. Wendt, G. R. & Collins, J. J. 3rd. Schistosomiasis as a disease of stem cells. *Curr. Opin. Genet. Dev.* **40**, 95–102. <https://doi.org/10.1016/j.gde.2016.06.010> (2016).
49. He, P. *et al.* The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* **583**, 760–767. <https://doi.org/10.1038/s41586-020-2536-x> (2020).
50. Howe, K. L. *et al.* WormBase 2016: Expanding to enable helminth genomic research. *Nucleic Acids Res.* **44**, D774–780. <https://doi.org/10.1093/nar/gkv1217> (2016).
51. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005> (2017).
52. Laverre, A., Tannier, E. & Necsulea, A. Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. *Genome Res.* **32**, 280–296. <https://doi.org/10.1101/gr.275901.121> (2022).
53. Hench, J. *et al.* The homeobox genes of *Caenorhabditis elegans* and insights into their spatio-temporal expression dynamics during embryogenesis. *PLoS One* **10**, e0126947. <https://doi.org/10.1371/journal.pone.0126947> (2015).
54. Lee, J., Chong, T. & Newmark, P. A. The esophageal gland mediates host immune evasion by the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 19299–19309. <https://doi.org/10.1073/pnas.2006553117> (2020).
55. Beall, M. J., McGonigle, S. & Pearce, E. J. Functional conservation of *Schistosoma mansoni* Smads in TGF-beta signaling. *Mol. Biochem. Parasitol.* **111**, 131–142. [https://doi.org/10.1016/s0166-6851\(00\)00307-8](https://doi.org/10.1016/s0166-6851(00)00307-8) (2000).
56. Loverde, P. T., Osman, A. & Hinck, A. *Schistosoma mansoni*: TGF-beta signaling pathways. *Exp. Parasitol.* **117**, 304–317. <https://doi.org/10.1016/j.exppara.2007.06.002> (2007).
57. Osman, A., Niles, E. G. & LoVerde, P. T. Expression of functional *Schistosoma mansoni* Smad4: Role in Erk-mediated transforming growth factor beta (TGF-beta) down-regulation. *J. Biol. Chem.* **279**, 6474–6486. <https://doi.org/10.1074/jbc.M310949200> (2004).
58. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: The power of many. *Nat. Methods* **17**, 11–14. <https://doi.org/10.1038/s41592-019-0691-5> (2020).
59. Crosnier, C. *et al.* Systematic screening of 96 *Schistosoma mansoni* cell-surface and secreted antigens does not identify any strongly protective vaccine candidates in a mouse model of infection. *Wellcome Open Res.* **4**, 159. <https://doi.org/10.12688/wellcomeopenres.15487.1> (2019).
60. Dinguirard, N., Heinemann, C. & Yoshino, T. P. Mass isolation and in vitro cultivation of intramolluscan stages of the human blood fluke *Schistosoma mansoni*. *J. Vis. Exp.* <https://doi.org/10.3791/56345> (2018).
61. Mann, V. H., Morales, M. E., Rinaldi, G. & Brindley, P. J. Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology* **137**, 451–462. <https://doi.org/10.1017/S0031182009991211> (2010).
62. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. <https://doi.org/10.1038/nmeth.2089> (2012).
63. Legland, D., Arganda-Carreras, I. & Andrey, P. MorphoLibJ: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* **32**, 3532–3534. <https://doi.org/10.1093/bioinformatics/btw413> (2016).
64. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587. <https://doi.org/10.1016/j.cell.2021.04.048> (2021).
65. Grant, C. E. & Bailey, T. L. XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv*. <https://doi.org/10.1101/2021.09.02.458722> (2021).
66. Davis, P. *et al.* WormBase in 2022—data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* <https://doi.org/10.1093/genetics/iyac003> (2022).
67. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
68. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. <https://doi.org/10.1186/s13059-019-1832-y> (2019).
69. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development* <https://doi.org/10.1242/dev.165753> (2018).
70. Rawlinson, K. A. *et al.* Daily rhythms in gene expression of the human parasite *Schistosoma mansoni*. *BMC Biol.* **19**, 255. <https://doi.org/10.1186/s12915-021-01189-9> (2021).

Acknowledgements

We thank our colleagues at the Wellcome Sanger Institute (WSI): Simon Clare, Cordelia Brandt, Catherine McCarthy, Katherine Harcourt and Lisa Seymour for assistance and technical support with animal infections

and maintenance of the *Schistosoma mansoni* life cycle; Geetha Sankaranarayanan and Magda Lotkowska for expert technical support; Sarah Buddenborg and Stephen R. Doyle for providing access to the latest versions of the *S. mansoni* genome assembly; David Goulding and Claire Cormie at the Electron and Advanced Light Microscopy facility for technical support with confocal microscopy; Nancy Holroyd, Mandy Sanders and Siobhan Austin-Guest for project and sequencing coordination. We also acknowledge Andrew Gillis from the Zoology Department, University of Cambridge for producing the image of a representative D5 mother sporocyst shown in Figure 1A. This research was funded in part by the Wellcome Trust [Grant numbers 206194 and 107475/Z/15/Z] and a UKRI Future Leaders Fellowship awarded to GR [Grant number MR/W013568/1].

Author contributions

Conceptualization: C.D., G.R.; Data Curation: C.D., T.A., Z.L.; Formal Analysis: C.D., T.A., Z.L., S.F., T.G.R.A., K.A.R., G.R.; Funding Acquisition: M.B.; Investigation: C.D., J.G., C.H., S.T., T.G.R.A., G.R.; Methodology Development: C.D., T.A., Z.L., S.F., T.G.R.A., K.R., G.R.; Project Administration: M.B., G.R.; Resources: C.D., T.A., G.R.; Software Programming: C.D., T.A., Z.L., S.F.; Supervision: M.B., G.R.; Validation: T.A., K.A.R., G.R.; Visualization: C.D., T.A., Z.L., S.F., G.R.; Writing—Original Draft Preparation: GR; Writing—Review & Editing Preparation: all authors.

Competing interests

Carmen Diaz Soria is currently employed at bit.bio, a cell coding company. All the experiments and analyses performed in this study occurred before Carmen Diaz Soria started to work at bit.bio. All other authors have no competing interest to declare.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-55790-3>.

Correspondence and requests for materials should be addressed to M.B. or G.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024